Exhibit 98

REDACTED VERSION FILED PURSUANT TO DKT. 350

UNITED STATES DISTRICT COURT FOR THE NORTHERN DISTRICT OF CALIFORNIA

DZ Reserve and Cain Maxwell (d/b/a Max Martialis), individually and on behalf of others similarly situated, Plaintiffs, v. FACEBOOK, INC., Defendant.	Case No.: 3:18-cv-04978

Expert Report of Charles D. Cowan, Ph.D.

December 22, 2020
San Antonio, TX

TABLE OF CONTENTS

I. Introduction

1. My name is Charles Cowan. I reside in San Antonio, Texas. I have been retained as a Statistics Expert by Plaintiffs' Counsel, Cohen Milstein Sellers & Toll PLLC, in the above referenced case.

2. The defendant in this action is Facebook, Inc., a California Corporation. It is my understanding that Plaintiffs and class members purchased online advertisements from Facebook. It is also my understanding that Plaintiffs allege that the Potential Reach Facebook displays to all class members on Facebook's Ads Manager, both initially and at the conclusion of any demographic targeting is inflated.

3. I was retained by Cohen Milstein Sellers & Toll PLLC, counsel for the Plaintiffs, to develop a methodology for estimation of the inflation of measures of Potential Reach provided by Facebook to the named Plaintiffs and each advertiser in the above referenced case. I rely on well accepted statistical procedures to conduct my research and to draw my conclusions. A more fulsome presentation of my assignment, my methods, and my resultant opinions follows in the remaining sections of my report.

II. Professional Qualifications and Compensation

4. I have over forty years of experience in statistical research and design. I received my Bachelor of Arts degree in Economics from the University of Michigan, my Master of Arts degree in Economics from the University of Michigan, and my doctorate in Mathematical Statistics from the George Washington University. I currently consult for numerous public and private sector entities on the design, implementation, and evaluation of research, and the synthesis of statistical and sampling techniques for measurement. My professional experience and academic tenure are included in my curriculum vitae, a true and correct copy of which is attached as Exhibit 1.

A. *Professional Experience*

5. I have designed large and complex research programs incorporating sampling. These include the Post Enumeration Program for the U.S. Census Bureau to evaluate the Decennial Census, the Economic Cash Recovery valuations conducted by the Resolution Trust Corporation ("RTC"), and evaluation studies conducted for the Department of Justice, the Department of Defense, and the Department of the Treasury.

6. From January 2002 to the present, I have been a Member of Analytic Focus LLC. My firm provides analytic services in the public and private sectors. My firm has multiple projects with the federal government involving audits and other review services. The firm helps businesses and nonprofits optimize their operations and estimate the economic impact of their activities. Included in the firm offerings are expert witness and consulting services in litigation. A list of cases in which I have given expert testimony during the previous four years is attached as Exhibit 2.

B. *Experience in Academia*

7. I have taught graduate and undergraduate courses in sampling theory, survey methods, statistics, and computer methods for analysis. I recently retired from my position as Professor in the School of Public Health at the University of Alabama at Birmingham.

8. I also served as an Associate Professor of Statistics at George Washington University from 1993 to 1998, and served as a Visiting Research Professor at the Survey Research Laboratory of the University of Illinois from 1983 to 1989.

C. *Publications*

9. I have co-authored two books: one on evaluation of survey and census methods, and one on econometric measures related to the welfare of the U.S. economy. I also have written numerous articles on statistical methods, sampling, rare and elusive population

research, and optimization techniques. A listing of these publications is included at pages 4 to 7 of my curriculum vitae, attached as Exhibit 1.

D. *Professional Societies*

10. I am a member of the American Statistical Association and have held memberships in other professional societies. My positions on professional committees are listed at page 3 of my C.V., attached as Exhibit 1.

E. *Compensation*

11. I am being compensated for my work on this engagement at the rate of $775 per hour for my time. The payment of my fees is not contingent on the opinions I express in connection with this engagement.

III. Assignment and Summary of Opinion

12. My task was to examine the numbers provided to advertisers by Facebook, which, according to Facebook, represent the number of people who could potentially see an advertisement that the advertiser is purchasing. Facebook describes these numbers as the "Potential Reach" of an advertisement, and so I use the same term throughout this report.

13. I have been asked to examine the Potential Reach numbers to determine if these numbers were inflated. Specifically, I have been asked to calculate with a reasonable degree of statistical certainly:

a. the inflation in the initial, default Potential Reach number shown to all US advertisers; and
b. the minimum Potential Reach number at which an advertiser, regardless of the targeting criteria they selected, would receive a Potential Reach inflated by 10% or more.

14. I find that even accounting for possible variation across different subgroups (target audiences generated using granular selection criteria), based on the global or

national numbers provided by Facebook, the likelihood that a Potential Reach number provided by Facebook is not inflated is essentially zero.

15. I find that because of the various ways that the Potential Reach may exceed the actual number of people who could potentially see an ad, that to a reasonable degree of statistical certainty:

a. the initial default Potential Reach provided to Plaintiffs and all advertisers in the U.S. is inflated by 44.5%;

b. if an advertiser receives a Potential Reach of 1000 or above, regardless of the targeting criteria used, their Potential Reach is inflated by at least 10%, and even for an advertiser receiving a Potential Reach below 1000, the likelihood of receiving a Potential Reach inflated by at least 10% is a near statistical certainty.

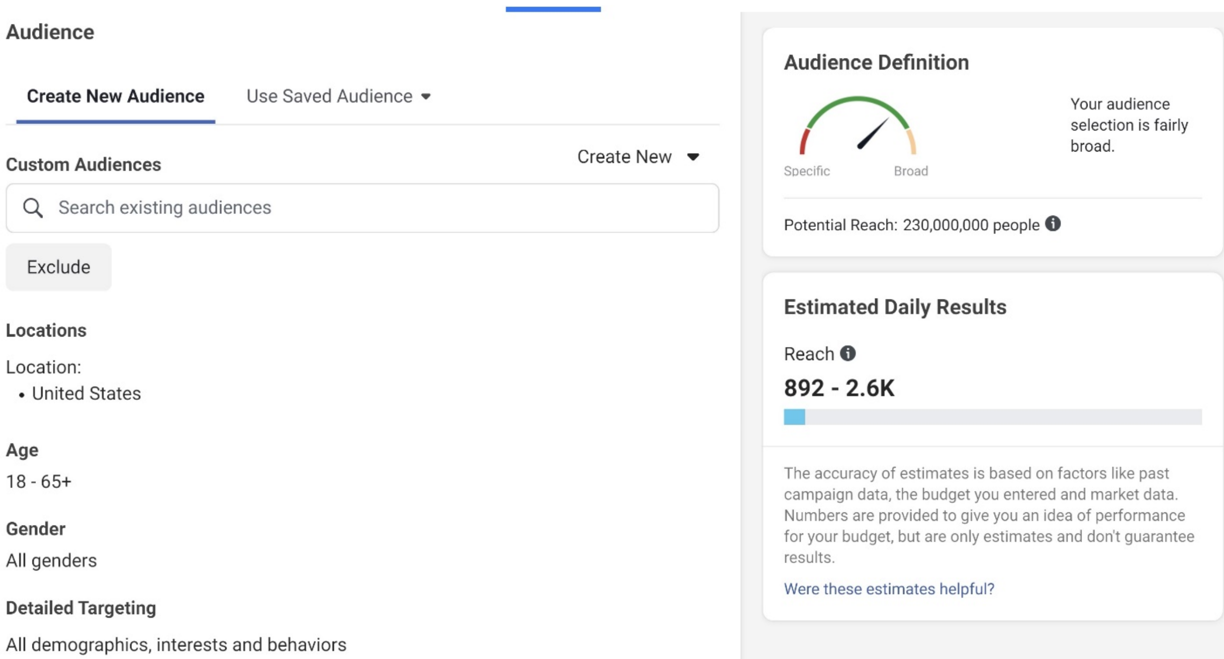
16. The methodology outlined herein applies equally to each named Plaintiff and to each advertiser individually; and I apply the same methodology to the Potential Reach received by the named Plaintiffs and to any advertiser.

IV. Case Background

17. It is my understanding from reading the pleadings and other materials in this case that Facebook sells advertisements to advertisers, which advertisers purchase through Facebook's Ads Manager interface ("Ads Manager"). On Ads Manager, advertisers can target their advertisement to users of different demographics. Using the demographic criteria on Ads Manager, advertisers can target people in various locations, age ranges, or genders. However, regardless of the demographic criteria used to target the advertisements, Ads Manager shows a graphic with the audience size based on the criteria that advertisers selected. Ads Manager displays a default Potential Reach number for the entire United States, which is shown in Figure 1 of the Third Amended Consolidated Complaint as "Potential Reach: 240,000,000 people."

18. As seen in Figure 1 below, as of December 12, 2020, the default Potential Reach for the United States read "Potential Reach: 230,000,000 people."

Figure 1



19. It is my further understanding from reading the pleadings in this case that plaintiffs allege that (a) the initial, default Potential Reach shown to all US advertisers is inflated; and (b) the Potential Reach displayed at the conclusion of demographic targeting (if any) is inflated.

20. I understand, that there have been public reports that Facebook's Potential Reach metric was inflated. For instance, in December 2016, the UK Advertising Standards Authority (UK ASA) complained to Facebook that Potential Reach was misleading. In 2017, various news publications, including AdNews, New York Times, Bloomberg, CNBC, and Business Insider, reported that Facebook claimed to be able to reach more people than possible, based on the findings of an analyst at the Pivotal Research Group. Likewise, in the fall of 2017, the Video Advertising Bureau (VAB) published a report that alleged that

Facebook's Potential Reach was inflated and exceeded the Census numbers in the US and other countries.

21. From reading the pleadings and documents produced in this case, I also understand that complaints about Potential Reach being misleading have been made since about September 2015 and that Facebook employees internally acknowledged that Potential Reach is inflated.

22. I understand, from reading the pleadings in this case, that Plaintiffs allege that the Potential Reach of Facebook advertisements is inflated, in part because the Potential Reach includes duplicate and fake accounts, and that Plaintiffs allege that Potential Reach is false and misleading because Facebook represents that Potential Reach counts people, but it really counts *accounts*.

V. Background on Analysis

23. My analysis is based on Facebook's own admitted inflation rates, as reflected in internal and public-facing documents, to which I applied widely accepted statistical principles. A list of the materials I relied upon in my analysis is attached to this report as Exhibit 3.

24. I chose to use information on sources of inflation in Potential Reach from Facebook's documents to be conservative in my assumptions. I believe Facebook's numbers are conservative compared to the inflation rates that are implied by comparisons to the Census. For instance, as stated in the Third Amended Complaint, in 2017 Facebook claimed that advertisers could potentially reach 100 million people ages 18-34 in the US when, according to the U.S. Census Bureau, there were only 74 million 18-34 year-olds in the US to begin with, and according to Pew Research only 80% of them were on

Facebook.[1] This, Plaintiffs alleged, resulted in a nationwide inflation rate of 64%.[2] Compared to that inflation rate, the inflation rates I calculated based on Facebook's numbers are very low.

25. Facebook's own internal documents are inconsistent about rates of Potential Reach inflation from various sources. I understand that Facebook has taken the position that there was no reliable data about Potential Reach inflation for situations in which advertisers selected a target audience using criteria more granular than just country and age[3]: for instance, a Facebook witnesses testified that, the model developed to remove one source of inflation (duplicate accounts) from the Potential Reach was never implemented because Facebook was unable to verify the model's accuracy across the many potential customized targeting demographics.[4] Moreover, I understand from the Expert Report of Dr. Atif Hashmi, who examined Facebook's source code, that Facebook utilized the same infrastructure—specifically, a sampling-based methodology—for calculating Potential Reach *regardless* of specific targeting criteria.[5]

26. Therefore, again to be conservative, and consistent with Facebook's own approach to calculating Potential Reach regardless of specific targeting criteria, I used the most robust, reliable data available from Facebook.

27. In typical statistical analysis, I would use a confidence level of 95% to draw conclusions from tests or research that I have conducted. This can be interpreted to mean that, were I to collect my data and conduct my analysis repeatedly in the same fashion and with the same sample size (though different samples), I would come to the same

---

[1] *See* Third Amended Consolidated Complaint (ECF No. 166) at ¶ 43.

[2] *Id.*

[3] *See, e.g.,* FB-SINGER-00274929; FB-SINGER-00169905; *see also* FB-SINGER-00025712; David Amsallem Deposition, at 69:17-70:12.

[4] *See, e.g.,* Yaron Fidler Deposition at 252:8-19, 256:1-257:20.

[5] Expert Report of Dr. Atif Hashmi at ¶¶ 40-42; *see also id.* at ¶¶ 34-39.

conclusions 95% of the time. In my professional experience running a quality assurance program for the Department of Justice, I tested for a 1 in 1000 and 1 in 10,000 error rate.

28. Here, I was asked to determine the probability of inflation of at least 10% when considering Potential Reach in a variety of situations. I found that the probabilities are well beyond the standards usually seen in litigation or typical statistical analysis. The confidence levels I found are what I would refer to as reasonable statistical certainty.

29. As explained above and herein, I did not have reliable data for the rates of inflation from all potential sources of inflation or every potential sub-demographic. But, I used accepted statistical techniques to determine the probabilities of minimum inflation notwithstanding the lack of more granular information regarding specific inflation levels of sub-demographics. I make the reasonable assumption that mechanisms that apply at the higher levels of targeting apply to more granular targeting; as I explain in Section VI.G, this assumption is reasonable in light of Facebook's own use of a sampling-based methodology, which presupposes at least a minimally haphazard distribution of each potential targeting specification. This allows me to use statistical methods to calculate the likelihood of inflation for any Potential Reach number generated by Facebook.

30. In particular, I utilized two standard tools in statistics to develop my analysis. I employed a Monte Carlo simulation, which allows the use of limited data provided by Facebook in discovery to determine the probabilities of inflation for a given Potential Reach, even when all the information about that population is not fully available. I also constructed likelihoods of inflation using a mean and variance computed using Facebooks own numbers for factors affecting inflation.

31. Monte Carlo simulations, and similar statistical techniques, are used routinely in real world applications and have been accepted by courts.[6] Indeed, the practice of using

---

[6] For instance, in *Lyondell Chemical Co. v. Occidental Chemical Corp.*, 608 F.3d 284 (5th Cir. 2010), which involved environmental cleanup costs, the Monte Carlo method successfully withstood a

limited available information to make broader observations across a larger population of data is a primary purpose of statistics. That is why these statistical methods are appropriate here. (And, conversely, if perfect data existed regarding inflation across all potential sub-demographics, a statistical analysis would not be necessary here.)

VI. OVERVIEW OF METHODOLOGY

32. In this section I describe the methodology I employ to calculate Potential Reach inflation based on the most robust and reliable data about sources of inflation available from Facebook's own documents. In Sections VI.A and VI.B, I assess the different sources of Potential Reach inflation—namely Fake accounts, SUMA accounts, Ineligible accounts, and Inactive accounts—their corresponding rates, as well as their distributions, based on information available from Facebook's own documents. In Section VI.C, I identify the formula for calculating Potential Reach inflation due to these four sources. In Section VI.E, I explain the range of overlap between the four sources of inflation that is possible and why the formula for Potential Reach inflation already assumes a natural overlap within that range. In Section VI.F, I calculate the range of Potential Reach inflation by country and age group, and in Section VI.G, I explain why that method applies to more granular targeting subgroups. In Section VI.H, I discuss why distribution of Potential Reach (and Potential Reach inflation) conforms to the normal distribution, and how that allows me to calculate to a statistical certainty the probability that Potential Reach above certain minimum thresholds is inflated by at least 10% and the inflation range for any given Potential Reach. Finally, in Section VI.I, I explain four methodologies for imputing missing

---

Daubert challenge, with the court concluding that the Monte Carlo method of statistical analysis, as employed to measure probability of various outcomes when reaching exact numerical result is impossible or infeasible and when data provides known range was reliable. *Id.* at 293.

Potential Reach values in order to identify ads with Potential Reach below any set threshold.

A. Sources of Inflation

33. Facebook's internal documents identify at least five independent sources of inflation, four of which I will address here, with the fifth addressed in Section VII.A and Appendix 1. Numbers in parentheses are derived from Facebook's documents. I will be using them in my analysis throughout this report.[7]

- **Fake** (5% Global): For the purposes of this report, these are defined as fake (i.e. not a person) or misidentified accounts that are actually associated with a business or some other non-person entity.[8] Facebook's SEC filings for 2015-2020 reflect that the number of Fake accounts has grown over the years. Estimates of Fake accounts conducted in 2014-2015 reflect a Fake account rate of less than 2%. Estimates of Fake accounts conducted in 2017 show Fake account rates ranging from 2% to 4%. Estimates of Fake accounts conducted in 2018 and 2019 show a Fake account rate of 5%. Because Facebook provided the most fulsome information about advertising sets purchased in 2018, I use the Fake account rate reflected in the 2018 10-K, which is 5%.

---

[7] As explained more fully below, I use distributional data contained within Facebook's own documents with respect to Ineligible (FB-SINGER-00170105) and SUMA (FB-SINGER-00426295) sources of inflation. For Fake and Inactive sources of inflation, however, I use a static number because Facebook did not provide reliable data concerning the distribution of inflation.

[8] This is consistent with Facebook's own definition. In its SEC disclosures, including the 2018 10-K, Facebook divides fake or false accounts into two categories: "(1) user-misclassified accounts, where users have created personal profiles for a business, organization, or non-human entity such as a pet (such entities are permitted on Facebook using a Page rather than a personal profile under our terms of service); and (2) undesirable accounts, which represent user profiles that we determine are intended to be used for purposes that violate our terms of service, such as spamming."

- **Ineligible** ([redacted] Global, [redacted] US): For the purposes of this report, this term refers to accounts that have not received an ad in the last 30 days, are not shown ads for some reason, or who are known to have an active ad blocker that keeps them from seeing an ad.[9] It is my understanding that Facebook itself acknowledged this discrepancy as a source of inflation (or, to use Facebook's terminology, "over-estimation"). FB-SINGER-00184892 identifies the global rate of this type of inflation as [redacted] FB-SINGER-00170105 identifies the US rate of this type of inflation as [redacted]
- **Inactive** (8.1% Global): For the purposes of this report, this term refers to accounts that are included in Potential Reach but are not included in Facebook's Monthly Active Users. They are best defined as the ratio of Potential Reach to Monthly Active Users (MAU). The Potential Reach cannot exceed the Monthly Active Users because the number of people who could potentially see an ad cannot be more than the number of active users in the first place. FB-SINGER-00258878 states that as of January 16, 2018, the Potential Reach for advertisers targeting US and Canada, as shown on Ads Manager, was 260 Million. But, Facebook's recent investor presentation, Facebook Q4 2019 Results, which can be found at https://s21.q4cdn.com/399680738/files/doc_financials/2019/q4/Q4-2019-Earnings-Presentation-_final.pdf shows that for Q4 of 2017, the number of Monthly Active Users in US & Canada was only 239 Million. FB-SINGER-00258878 also echoes that at '79: "NA [North America] reported MAP [Monthly Active People] for Q3 2017: 239 M". If there were 239 million Monthly Active Users but, at the same time, 260 million "people" in the Potential Reach, then the Potential Reach contained 8.1% more "people" than did the measure of Monthly Active Users. This

[9] *See* FB-SINGER-00184892.

report will refer to these 8.1% of accounts as "Inactive" solely as a shorthand and for the sake of simplicity.

- **SUMA** ([redacted] Global, [redacted] US): These are defined as Single User, Multiple Accounts (SUMA) [10], i.e., multiple accounts on Facebook (excluding any additional accounts they may hold on Instagram) that all refer back to a single person. In documents provided by Facebook, there are a number of different values presented for SUMA. (I address this further in Appendix 5). I rely on FB-SINGER-00426295, a table provided in discovery that allows me to compute SUMA for 240 countries, up to 11 age groups in each country. I sum the Potential Reach and Adjusted for SUMA Potential Reach numbers in this table across all countries and all age groups to obtain a global value for SUMA ratio in Appendix 2A (which contains the Potential Reach inflation rates calculated by different country and age groups). From Appendix 2A, I calculate the global SUMA ratio to be [redacted] and the U.S. SUMA ratio to be [redacted] see also Appendix 2B).

34. Though I rely on the above rates for the four sources of inflation throughout this report, I perform additional sensitivity analyses in Appendix 11.

35. In Section VII.A and Appendix 1 to this report, I add a fifth source of inflation: the proportion of people who have both Facebook and Instagram accounts that are not deduplicated from one another. As noted below, this source of inflation is relevant for purposes of determining the inflation of Facebook's default initial Potential Reach displayed on Ads Manager. This source of inflation is also relevant for purposes of determining the inflation of any individual Potential Reach where the advertiser selected

---

[10] *See, e.g.,* FB-SINGER-00301080.

a placement on Facebook and Instagram, which can be determined by based on Facebook's internal records.[11]

B. Distributions of Inflation Sources

36. It is my understanding Facebook conducted limited analysis of the distribution of inflation sources, but that the most reliable analysis was at the age and country level.[12] FB-SINGER-00426295 shows the distribution of SUMA rates by country and age, as well as the inflation factor. I include further discussion of the SUMA rates from Facebook's documents in Appendix 5.

37. For Ineligible accounts, there is one source[13] that lists the proportion of Ineligible accounts for 30 countries, with an average Ineligible rate of [redacted] In addition, another source[14] states that the global Ineligible rate is [redacted] Using a regression, I can relate the Ineligible rate to the SUMA rate for the countries provided in Facebook's documents[15]. The countries are: Russian Federation, Japan, Indonesia, India, Viet Nam, Korea, Republic of, Thailand, Spain, Hong Kong, Turkey, France, Mexico, Germany, Taiwan, Israel, Singapore, Poland, United States, Netherlands, Belgium, Argentina, Canada, Australia, United Kingdom, Malaysia, Brazil, Italy, Sweden, Denmark, and Norway.

38. In creating a plot of Ineligible rates by SUMA rates for all countries and age groups, there is an apparent relationship that shows that the Eligible rate (one minus the Ineligible rate) increases as SUMA increases. I undertook to fit a regression between eligible rates and SUMA rates to determine if the visual relationship is confirmed by the

---

[11] *See* FB-SINGER-035, demographic targeting data for ads that ran in 2018.
[12] *See* FB-SINGER-00025712; David Amsallem Deposition, at 69:17-70:12.
[13] FB-SINGER-00170105.
[14] FB-SINGER-00184892.
[15] FB-SINGER-00426295. I use least squares regression with indicator variables for each age group in the data set, as well as indicator variables for estimates from Russia, South Korea, and Japan.

regression. The reason for using a regression is to take advantage of this relationship and to allow some Eligible account rates to be higher, some lower, to reflect the variability and trend between SUMA and Eligible accounts. The alternative would be to simply apply an Ineligible rate of [redacted] all countries with no variability.

Chart 1: Predicted and Actual Eligible Rates by SUMA for 30 Countries – All Observations (Outliers in Regression Excluded from Chart)

39. Using a regression, I forecast the Eligible rate for every other country except the 30 countries provided in the discovery for Facebook. For the 30 countries I used exactly the values provided by Facebook even though some of them seem extreme. For the remaining 210+ countries and age groups I use the values estimated from the regression. More details on the regression are available in Appendices 7 and 10. It is important to note that the values that are input into the regression from the 30 countries selected by

Facebook have lower ineligible rates, averaging about [redacted] ineligible, relative to the [redacted] Facebook states is the global rate. This means that my regressions are exceptionally conservative as the average prediction from the regression would only be [redacted] ineligible, not [redacted]

40. As described above, using Facebook's documents I was able to calculate the distribution rates for SUMA and Ineligible accounts at the country and age level. However, I had no such distributions for the rates of Fake and Inactive accounts. Therefore, I used a global rate of Fake Accounts of 5% from Facebook's SEC filings and a rate for Inactive accounts of 8.1% from Facebook's documents. Should more granular information about these two rates become available at the country and age level, I can easily use exactly the same methods to recompute the estimated Potential Reach inflation rate.

C. Computation of Inflation Using Any One of Four Sources

41. Combining the sources of inflation outlines above into a single formula, I estimate the degree to which the Potential Reach numbers provided by Facebook exceed the actual number of people who could potentially see the ad. I refer to this excess as "inflation" through the remainder of this report. Inflation is simply the ratio of Potential Reach to the number of actual people who could potentially see the ad.

42. An inflation factor can be computed using the data that Facebook has provided. Inflation is defined to be:

$$Inflation = \frac{Potential\ Reach}{Actual\ People\ Who\ Could\ Potentially\ See\ an\ Ad}$$

43. If there were only one source of inflation, this computation would be very simple. For example, suppose that there were only Fake accounts, that there was only one Fake

Account created, and there were no Ineligible accounts, no Inactive accounts, and no multiple accounts for a single user (i.e. no SUMA)[16]. Then inflation would be defined as:

$$Inflation = \frac{Potential\ Reach}{Actual\ People} = \frac{Real + Fake\ Accounts}{Real\ Accounts} = 1 + \frac{Fake\ Accounts}{Real\ Accounts}$$

44. If the rate of Fake accounts out of all accounts is 5%, as Facebook reported to the SEC, then inflation due solely to Fake accounts would take the value 1.05. The same computation can be performed for other sources of inflation. In each case, the ratio is simply the inflation from that one cause.

45. However, with multiple sources of inflation the computation becomes more complex. The difficulty is in ensuring that inflation from multiple sources is correctly determined and not duplicative.

D. Computation of Inflation Using Four Sources Simultaneously

46. In Section VI.A I defined Fake, SUMA, Ineligible, and Inactive accounts. To track people who could be shown an advertisement, I need to re-express these definitions by their opposites:

- **Genuine** (95% Global)
- **Eligible** [redacted] Global, [redacted] US)
- **Active** (91.9% Global)
- **SUSA (Single User, Single Account)** [redacted] Global, [redacted] US)

47. The SUMA ratio, based on a global SUMA inflation rate of [redacted] would be expressed as

[16] In this hypothetical situation, once Fake accounts are removed, the number of accounts equals the number of people.

- **SUMA ratio** = Total of Multiple Accounts \ Number of People = [redacted]

48. Combining these is simple arithmetic. I start with the Potential Reach number. First, I multiply Potential Reach by the Genuine rate to obtain the number of Genuine accounts. Second, I multiply the number of Genuine accounts by the proportion of Eligible accounts to obtain Genuine & Eligible accounts. Third, I multiply this Genuine & Eligible number by the proportion of Active accounts to obtain Genuine & Eligible & Active accounts. Fourth, I divide the number of Genuine & Eligible & Active *accounts* by the SUMA rate to obtain the number of *people* with Genuine & Eligible & Active accounts. In other words: the number of active, non-duplicate accounts on Facebook that belong to real people, and are eligible to see ads, is the same as the number of people who could see ads on Facebook.

49. The arithmetic involves four simple steps, summarized in Table 1 below.

Table 1: Steps to Compute Inflation

	Step in Computation	**Symbols**	**Global Values**
	Potential Reach		
1	*x Genuine Rate*	(1-f)	95%
	= Actual Accounts		
2	*x Eligible Rate*	(1-g)	[redacted]
	= Eligible Actual Accounts		
3	*x Active Account Rate*	(1-v)	91.92%
	= Eligible Active Actual Accounts		
4	*/ SUMA Rate*	SUMA	[redacted]
	= Eligible Active People		

50. The order in which I perform these operations doesn't matter[17]. Using the steps in Table 1, I get:

[17] Because of the simple rule, a*b = b*a

$$Potential\ Reach * (1 - f) * (1 - g) * (1 - v) \div SUMA = People\ Who\ May\ See\ the\ Ad$$

51. I rearrange these terms to get the formula for Potential Reach inflation rate that I will rely on throughout this report.

$$\boldsymbol{Inflation\ Rate = \frac{Potential\ Reach}{People\ Who\ May\ See\ the\ Ad} = \frac{SUMA}{(1-f)(1-g)(1-v)}}$$

52. I adopt a few assumptions related to the process of counting these accounts, namely:

i. Fake accounts count as one each; they do not have duplicate Facebook accounts.

ii. SUMA rate for each individual Facebook account is enumerated as 1, 2, 3, …; while most accounts will have a single user, some have more than one. If no one had multiple accounts, SUMA rate would be 1.0.

53. The SUMA rate is the ratio of accounts to people for Facebook; it is calculated as:

$$\begin{aligned} SUMA = [\, 1 * (\%\ of\ people\ with\ one\ account\ in\ FB) \\ + 2 * (\%\ of\ people\ with\ 2\ accounts) \\ + 3 * (\%\ of\ people\ with\ 3\ accounts) + \cdots] * People \end{aligned}$$

where the % of people with 1, 2, 3, … accounts sums to 100%.

54. While SUMA refers to duplicate accounts for the same user on Facebook, Facebook's own documents show that users can have 2, 3, 4, 5, or more duplicate accounts.[18] But, for this report, I only have the SUMA rates reported by Facebook, but not

[18] *See, e.g.,* FB-SINGER-00314173. I discuss this further in Appendix 5.

the details of whether or how those SUMA rates account for users who have more than two duplicate accounts. Therefore, I make the conservative assumption that Facebook users have either one account (SUSA) or two duplicate accounts (SUMA). This results in using exactly the SUMA rates provided by Facebook while allowing computation of a process where some Facebook users have opened more than one account.

E. Overlap Between Sources of Inflation

55. I assume that the four sources of inflation are independent of one another, and thus assume a reasonable overlap between them. In this section, I will discuss why that assumption is reasonable, given the maximum and minimum overlap that is possible for the four inflation sources. The calculations for the Tables below are also in Appendix 4.

56. The analyses performed by Facebook are performed only one at a time, so there is no indication that the factors are related based on any data produced by Facebook. Further, I have seen no documents suggesting that there may be a correlation between any of the inflation sources I have analyzed.

57. Assuming the four sources of inflation are independent, and using the rates of Fake, Inactive, Ineligible, and SUMA accounts as stated in Section VI.A, I can compute the likelihood of the presence or absence of all four sources combined. This computation depends on a simple elementary tenet of probability theory: the probability two events being independent is computed as the probability of the first event times the probability of the second. This is just the probability of A and B both occurring computed as P(A,B) = P(A)*P(B). I do this for all the sources of inflation. The results are in Table 3 below.

Table 3: Probability of Four Sources of Inflation Occurring or Not Occurring, Assuming Independence Between the Four Sources

% SUSA,[redacted], One Account

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	**66.80%**	11.19%	78.0%
No	**5.87%**	**0.98%**	6.9%
Total	72.7%	12.2%	84.8%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	3.52%	0.59%	4.1%
No	0.31%	0.05%	0.4%
Total	3.8%	0.64%	4.5%

% SUMA,[redacted] Accounts

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	7.99%	1.34%	9.3%
No	0.70%	0.12%	0.8%
Total	8.7%	1.5%	10.2%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	0.42%	0.07%	0.49%
No	0.04%	0.01%	0.04%
Total	0.46%	0.08%	0.53%

58. With this set of outcomes, the overlap that indicates people who may see an advertisement as a percent of the number presented as Potential Reach is only 66.8%. Based on the values in Section VI.D, this would be computed as Active (.9191) * Eligible [redacted] * Genuine (.95) * SUSA[redacted] = .668 or 66.8%, assuming those factors are independent. This means that assuming all four sources of inflation are not correlated, for any given account there is a 66.8% chance that it is completely free form any sources of inflation – i.e., is Eligible, Active, Genuine, and SUSA. This assumes a natural overlap between the sources of inflation, based on a random correlation between them.

59. Though I have seen no documents suggesting this, I could instead assume that the inflation sources are correlated. I accept the rates for all four sources of inflation based on Facebook's documents as fixed—that is, these four values are preserved even when an overlap is considered—then I can readily compute the maximum and minimum overlap that could occur. These values are computed by considering the lowest likelihood that any given account is Eligible, Active, SUSA, and Genuine (66.8% - 0.98% = 65.82%), and then the highest likelihood that any given account Eligible, Active, SUSA, and Genuine (66.8% + 5.87% = 72.67%). This computation is shown in Table 4 below.

Table 4: Likelihood Any Account Represents a Person (i.e. No Inflation) in Potential Reach Assuming Minimum and Maximum Overlap in Sources of Inflation

Minimum Overlap

Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**65.82%**	12.18%	78.0%
No	6.85%	0%	6.9%
Total	72.7%	12.2%	84.8%

Maximum Overlap

Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**72.67%**	5.32%	78.0%
No	0%	6.85%	6.9%
Total	72.7%	12.2%	84.8%

60. Table 4 shows that the minimum overlap is computed by assuming that there are no Ineligible and Inactive accounts (and also no SUMA or Fake accounts). To preserve the totals in the table, the proportion of Eligible and Active accounts has to decline by the same proportion as the proportion of Inactive and Ineligible accounts (note the value of zero percent in the left table in Table 4). The maximum overlap is computed as the maximum proportion that can be Eligible, Active, Genuine, and SUSA accounts, which is computed by assuming there are no accounts that are jointly Eligible and Inactive.

61. This represents a range of inflation of approximately 1.36 to 1.50, or 36% to 50%. I calculated these ranges as follows:

- Inflation assuming minimum overlap = SUMA ratio [redacted] (SUSA ([redacted] / Likelihood under Maximum Overlap (.7267)) = 1.360.
- Inflation assuming maximum overlap = SUMA ratio [redacted] (SUSA ([redacted]) / Likelihood under Minimum Overlap (.6582)) = 1.502.

62. This means that, taking into account the possible range of overlap between all four sources of inflation, the lowest likelihood that any given account corresponds to a person (i.e. is Genuine, Active, Eligible, and not SUMA) is 65.8% and the inflation rate (which is inversely related to the overlap) in that case is 50%. Likewise, the highest likelihood that any given account corresponds to a person (i.e. is Genuine, Active, Eligible, and not SUMA) is 72.7%. and the inflation in that case is 36%.

63. The overlap I use in my computations, 66.8%, is the likelihood that any given account is not inflated when all four sources are independent. Table 4 shows that when the rates of the four inflation sources are preserved, the range of permissible values for this percentage is only 65.8% to 72.7%. So, 66.8% is a reasonable number within the narrow range.

64. The corresponding average global rate of inflation is SUMA ratio [redacted] (Active (.9192) x Eligible [redacted] Genuine (.95)) = 1.48, or 48%.[19]

F. Range of Inflation Across Country and Age Groups

65. Using the limited reliable information from Facebook, I developed estimates of Potential Reach inflation by country and by age group.

---

[19] This does not include inflation caused by users who have a Facebook and Instagram account that are not deduplicated from one another. If this source were included, the Potential Reach inflation rate would be higher.

66. I calculate this distribution based on the range of SUMA rates from FB-SINGER-00426295 by country and age, the range of Ineligible rates described in Section VI.B (which used FB-SINGER-00170105 as the starting point), a global 5% rate for Fake accounts and a global 8.1% rate for Inactive accounts. As discussed in Section VI.E, my calculations of inflation by country and age assume a reasonable overlap between the four sources of inflation.[20]

67. The formula in which I utilize the various SUMA and Ineligible rates (by country and age) and the static Fake and Inactive rates is, again, as noted in Section VI.D:

$$Inflation\ Rate = \frac{Potential\ Reach}{People\ Who\ May\ See\ the\ Ad} = \frac{SUMA}{(1-f)(1-g)(1-v)}$$

68. The full distribution by country and age group is available in Appendix 2A.

G. Range of Inflation Across Non-Age/Country Sub-Demographics

69. It is my understanding that in addition to country and age, numerous other more granular targeting parameters exist.[21] These targeting parameters include items such as interests that Facebook believes people have, such as yoga, movies, books, and other activities or items discussed on Facebook, and other factors that an advertiser may specify.

70. The data Facebook produced in discovery, however, does not offer any reliable information regarding the degree of inflation at these more specific targeting parameters. As discussed above in Section V, to the extent it exists and was produced, Facebook's most reliable data about inflation in Potential Reach is at the country and age level.[22] As

---

[20] As discussed in depth in Section VII.A.1, there is a fifth source of inflation: duplicate cross-platform accounts on Facebook and Instagram. Further, just as there are duplicate accounts on Facebook only, called SUMA, there are documents indicating that there are duplicate accounts on Instagram only, however I do not have the rates for them. To be conservative, I do not account for either the cross-platform duplicates or the duplicate Instagram-only accounts in Appendix 2A.

[21] *See, e.g.,* FB-SINGER-00426324, FB-SINGER-00304046, FB-SINGER-00304054, FB-SINGER-00304060, FB-SINGER-00426302, FB-SINGER-00426314, FB-SINGER-00426321.

[22] *See, e.g.,* FB-SINGER-00274929.

noted above, Facebook disavowed the accuracy and reliability of any analyses of inflation done at a more granular level than country and age.[23] Moreover, I understand from the Expert Report of Dr. Atif Hashmi, who examined Facebook's source code, Facebook utilized the same infrastructure—specifically, a sampling-based methodology—for calculating Potential Reach *regardless* of specific targeting criteria.[24]

71. In light of the above, the underlying principle I rely on is that every possible targeting subset, such as interests and narrow geo-targeting, consists of persons who have an age group and country location. More narrowly targeted audiences—such as male yoga-lovers ages 18-34 who live in San Francisco—are just subgroups within country and age groups. Because Facebook has provided SUMA rates for every country and age group and because I can estimate the Eligible rates for each country and age group using a regression, I can also calculate the probable inflation for each of these subgroups using the information available on the country-and age group-level by making the reasonable assumption that the mechanisms that apply at the country and age group level apply at lower levels within those subgroups.

72. The assumption that mechanisms that apply at the higher levels of targeting apply to more granular targeting is a reasonable assumption in light of Facebook's own use of statistical sampling to generate Potential Reach numbers themselves. Facebook does not "count" every account to generate its Potential Reach; rather, it derives each Potential Reach number based on a sampling methodology that uses a small subset of accounts, which are then extrapolated to generate much larger Potential Reach numbers.[25] In short, Facebook's Potential Reach is a statistically generated number.

[23] *See, e.g.,* Yaron Fidler Deposition, at 252:8-19, 256:1-257:20.

[24] Expert Report of Dr. Atif Hashmi at ¶¶ 40-42; *see also id.* at ¶¶ 34-39.

[25] Expert Report of Dr. Atif Hashmi at ¶¶ 26-33.

73. Facebook's sampling-based methodology presupposes at least a minimally haphazard distribution of each potential targeting specification. Facebook uses a similar sampling methodology to report its Fake and SUMA rates in its SEC filings, which also assumes at least a minimally haphazard distribution of Fake and SUMA accounts throughout its platform. If the distribution of targeting criteria or inflations sources such as fake and duplicate accounts were not even minimally uniform, but were instead highly concentrated in specific sub-demographics, then the relatively small samples utilized by Facebook every day in its own business would not be feasible. In sum, my distributional assumption is consistent with the assumption Facebook itself utilizes, and it is therefore reasonable to apply that assumption to each of the named Plaintiffs' Potential Reach numbers, and to any individual advertiser.

74. In addition, my methodology anticipates substantial variations in the distribution of inflation sources across demographic sub-groups. The Monte Carlo method considers the interaction of the four sources of inflation simultaneously and how they jointly lead to Potential Reach inflation. If the model underlying the four sources operating jointly led to a clustering of Potential Reach values with zero inflation, it would have developed in the 10,000 iterations I used (see Section VII.B.1). If there were systematic clusters of purity with no inflation, Facebook never analyzed it in its internal documents. Nor did Facebook design its sampling procedures to capitalize on this bifurcation, and when extrapolating results from samples it never employed post-stratification to account for these pockets of non-inflation. I can only conclude that, from the methods used by Facebook for measuring SUMA and other inflation sources, that Facebook does not know of or cannot identify pockets where the basic factors that lead to inflation do not operate. In addition, as described in Appendix 11, I have conducted several sensitivity analyses demonstrating that my calculations are robust even if various inflation sources are completely removed. For these reasons, and as explained more fully below, I believe that my distributional

assumption is reasonable and that my analysis utilizes all available reliable data to calculate the inflation for the named Plaintiffs and any individual advertiser.

H. Distribution of Potential Reach and Probabilities of Inflation

75. Using the available information, I can establish that Potential Reach inflation is distributed according to the normal distribution (also known as the Gaussian distribution). Establishing the distribution of Potential Reach inflation, will in turn, enable me to determine (a) the threshold Potential Reach value beyond which it is a statistical certainty that inflation is at least 10%, as well as (b) the inflation range for any given Potential Reach value.

1. Potential Reach Is a Summation of Indicator Variables

76. Calculations of Potential Reach inflation are a mean value, applied to a group of accounts. Potential Reach is (supposed to be) a count of accounts. If there are 10,000 individual accounts, Facebook adds up a value of "1" for each individual account, 10,000 times. In doing so, Facebook tacitly assumes that every account is Active, Eligible, Genuine, and SUSA. This tallying of Potential Reach is not unlike the Decennial Census of Population and Housing done by the Census Bureau, in which each person is ostensibly counted once.

77. However, Facebook's own documents show that this count of accounts (tallying Potential Reach) is severely flawed due to the various sources of inflation. The issue of Potential Reach inflation is not unlike the issue of undercounting and overcounting individuals that the U.S. Census Bureau has faced, including during my time at the Bureau.

78. Statisticians deal with undercounting or overcounting by summing "indicator variables", which take on the value one or zero, depending on the indication, e.g., whether the account is Inactive or Active. A different indicator variable would take on the value zero or one depending on whether the account is Ineligible or Eligible, and so on.

79. Thus, a proper Potential Reach calculation would be a sum of a set of indicator variables. For any one account, in a group of accounts, the indicator variable is a random outcome. Summing these, one obtains a total of these random outcomes. The overall sum for a target audience group the advertiser specifies would be the number of people that are likely to see an advertisement. Potential Reach, as shown to the advertiser, divided by this total, is thus the mean inflation.

80. Summing indicator variables is a commonly accepted methodology in litigation. Take for instance, a case involving diminished property value due to environmental disamenities (like groundwater contamination), where claims of diminished value are examined by comparing housing values within the contaminated area to housing values outside the contaminated area. This is typically done through the use of a regression, where the sales price of a house is predicted based on factors like square footage and age, plus a single variable coded as a one or a zero, depending on whether a house in is within or outside of the affected area. The regression in that case produces a mean value for property value loss, which is the coefficient in the regression on this indicator variable. If the mean loss is statistically significant, then there is a loss for the area that is exposed to the contamination. The more houses in the analysis, the better the reliability on this mean estimate of loss.[26] Regression is also commonly used in Federal programs to study

[26] One can argue that the approach of summing indicator variable would not apply because each account has a different probability associated with being Active, Eligible, Genuine, and SUSA for each person. But Facebook itself uses averages when investigating the accuracy of Potential Reach. Because Facebook is able to compute averages and apply them at high levels to countries and age groups, it is reasonable and appropriate to apply a similar assumption that a similar mechanism exists that applies at more granular levels of targeting, to the subgroups. Just as each house in a neighborhood affected by groundwater contamination may have a different individual loss, these individual losses sum to an average loss which is tested by the regression. It is the *sum* of losses across all households that is relevant to whether a neighborhood is contaminated. In the same way, it is the total prevalence of Inactive, Ineligible, Fake, and SUMA accounts that affects Potential Reach, which is a sum, not a single unique number.

how patients respond to new drugs or treatments, holding constant other factors like body mass, age, family history, and other factors. One example is reaction to new drugs for treating heart attacks, since it is well known that the likelihood of a heart attack is a function of these variables, as well as whether the patient smokes, is exposed to toxic chemicals or air pollution, and other environmental factors.

81. My approach to estimating the range of Potential Reach inflation in this case is as follows. I will compute a mean inflation, then apply it to each advertisement, but with the understanding that the mean has some variability—it is, after all, a statistical calculation. I will also compute the range of potential outcomes for a Potential Reach of a particular size and show the range around the mean.

2. Sums of Indicator Variables Tend to Be Normally Distributed, So Potential Reach Will Be Too

82. According to the Central Limit Theorem[27], which is a foundational theorem in statistics and probability, sums of random variables tend to be normally distributed as the size of the group being summed grows. Since Potential Reach is properly expressed as the sum of random outcomes/indicator variables, pursuant to the Central Limit Theorem I can reasonably assume that the distribution of Potential Reach, and its inflation, will converge to the normal distribution. A second related theorem[28] allows me to conclude

[27] Kendall, Maurice and Stuart, Alan, *The Advanced Theory of Statistics*, Macmillan, New York, Vols. 1 and 2.

[28] Jack Hayya, Donald Armstrong, Nicolas Gressis, (1975). "A Note on the Ratio of Two Normally Distributed Variables." *Management Science* 21(11):1338-1341, available at: https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.21.11.1338. *See also* T. Pham-Gia, N. Turkkan & E. Marchand (2006). "Density of the Ratio of Two Normal Random Variables and Applications." *Communications in Statistics - Theory and Methods*, 35:9, 1569-1591, DOI: 10.1080/03610920600683689, available at: https://www.researchgate.net/publication/233479623_Density_of_the_Ratio_of_Two_Normal_Random_Variables_and_Applications; Hinkley, D. (1969). "On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 56(3), 635-639. doi:10.2307/2334671, available at: http://static.stevereads.com/papers_to_read/on_the_ratio_of_two_correlated_normal_random_var

that, under general conditions, the ratio of one normal variable to another normal variable is generally also random – thus if the sums in the numerator and denominator are normal, so is the ratio. This is helpful in that it enables me to compute a distribution to allow testing where the ratio 1.10 (the 10% inflation) falls. Put another way, I can compute the likelihood of observing a 10% inflation or lower inflation using the normal distribution, given the distributions I compute using Facebook's rates for the sources of inflation.

83. Based on the above, in Section VII.B, I will show the probability that any given Potential Reach above a minimum threshold is inflated by at least 10% by (1) running a Monte Carlo simulation using Facebook's rates for all four sources of inflation which will confirm a normal distribution, and also (2) calculating the variance of the distribution, as well using a formula derived to express the variance of the Potential Reach inflation rate.

I. Identifying Potential Reach Numbers Below Any Set Threshold

84. Part of my assignment is to determine the Potential Reach threshold beyond which Facebook's Potential Reach is inflated by at least 10% to near statistical certainty. In particular, I calculated the probabilities that the Potential Reach is inflated by at least 10% for Potential Reach above 250, 500, 1000, and 1100. Conversely, I can also identify the advertisements that fall below any set Potential Reach threshold, including that of 250, 500, 1000, or 1100.

---

iables.pdf; Hinkley, D. (1970). "Correction: On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 57(3), 683-683. doi:10.2307/2334796, available at: http://static.stevereads.com/papers_to_read/correction_on_the_ratio_of_two_correlated_normal_random_variables.pdf; George Marsaglia (April 1964). "Ratios of Normal Variables and Ratios of Sums of Uniform Variables." *Defense Technical Information Center,* available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/600972.pdf; Marsaglia, G. (2006)." Ratios of Normal Variables. *Journal of Statistical Software*. 16(4), 1 - 10. doi:10.18637. Available at: https://www.jstatsoft.org/article/view/v016i04

85. It is my understanding that Facebook produced transactional-level data regarding all advertisements on Facebook during the August 2014 – April 2019 portion of the class period in a data set called all_ads_details, FB-SINGER-026. Facebook also produced a database, FB-SINGER-035, which details targeting criteria used in ads run during 2018 only.

86. The all_ads_details data set provides information at the ad_id (ad) level, and also links ad_ids to their corresponding ad sets (campaign_id), groups of ad sets (campaing_group_id), and advertiser's account number (anon_account_id). The all_ads_details data set includes, among other characteristics, audience_size, lifetime_legal_revenue, lifetime_legal_impressions, ad_create_date. The Potential Reach field in Facebook's data set is called audience_size.[29]

87. The ad targeting details database also provides information at the ad_id level, but the corresponding campaign_id, campaing_group_id, and anon_account_id are not provided. Instead, the ad targeting details data set contains the country, age, gender, and other targeting criteria.

88. However, Facebook did not provide a Potential Reach number for each ad set purchased during the August 2014-April 2019 period. The all_ad details_dataset includes [redacted] records, but only [redacted] have positive lifetime_legal_revenue. Of these [redacted] records, Facebook only provided Potential Reach value for [redacted] 66%) records and failed to do so for the remaining [redacted] (34%). Moreover, many ad_ids in the ad targeting details data base are missing values for country, age, or gender.

89. Based on the data provided thus far, however, I can preliminarily determine that a relatively small number of ad sets have a Potential Reach below 250, 500, 1000 or 1100. Based on my review of the ads that ran for which Potential Reach records were actually

---

[29] *See* Gerardo Zaragoza Deposition at 87:8-16; *see also* FB-SINGER-00313499 at '3516.

provided[30], only 1.3% of the ad_ids (ads) have a Potential Reach of below 250, only 1.9% of the ads have a Potential Reach below 500, only 4.2% of the ads have a Potential Reach below 1000, and only about 4.4% of the ads have a Potential Reach below 1100.

90. Notwithstanding these small percentages, I can use accepted statistical principles to identify all the ads that received a Potential Reach below any given threshold, including 250, 500, 1000, or 1100.

91. First, I can use alternative estimation methods to impute Potential Reach values to identify ads with Potential Reach below a certain threshold. After I impute the missing Potential Reach numbers, I will be able to classify ads by range of Potential Reach, including identifying set of ads with Potential Reach below any given threshold.

92. First, the most fundamental method is aggregating the Potential Reach averages. I can impute the Potential Reach values for ads by looking at the Potential Reach values in ad sets, groups of ad sets, and so on. There is a high correlation—at least .90—among audience_size values of ad_ids within either campaign_group_id or campaign_id. Therefore, the average audience_size of ad_ids within either of these groups is a good estimator of a missing audience_size within the same group. For instance, consider a campaign_id with four ad_ids where one of the ad_ids has a missing audience_size value. This missing value can be imputed with the average audience_size of the other three

---

[30] In July 2020, Facebook produced a data set called all_ads_details, FB-SINGER-026, that contained various transactional level information (including Potential Reach) about ads on Facebook for the period of August 2014-April 2019. At approximately 2pm ET on Saturday, December 19, 2020, Facebook informed Plaintiffs' counsel via email that FB-SINGER-026 was missing all the data for October 2016, and transmitted additional file with the October 2016 data. Because the expert disclosures deadline was Tuesday, December 22, 2020, I did not have an opportunity to review and analyze the October 2016 data and do not consider it in my report. Based on Facebook's representation regarding the limited amount of missing data, I do not believe the updated version of the dataset will alter my analysis. I will, however, review the recently produced data set, and reserve the right to amend my analysis if appropriate.

ad_ids—or the average audience_size of all ad_ids within the same campaign_group_id as the ad_id at issue.

93. There are ad_ids for which the average audience_size is only available at the account id level. In such cases, I will use this average but in a different estimation method. In addition to using an average audience_size at campaign_id or campaign_group_id level for that account id, I could better tailor the audience_size estimate of an ad_id by using information specific to it from the all_ad_details database. For example, I could use the year and month when an ad was created (ad_create_date), its lifetime_legal_revenue, its lifetime_legal_impressions, etc.

94. Second, if there are still missing audience_size values, I can process all the information above to impute missing audience_size values through a regression model. Using the variables mentioned above, I could use the following regression model to impute missing audience_size values:

$$PR = \alpha + \beta_1\, avg\ PR + \beta_2\, year + \beta_3\, month + \beta_4\, revenue + \beta_5\, impressions$$

PR = ad_id's audience_size

avg PR = average audience_size of all ad_ids within the same campaign_id (or campaign_group_id) as the referred ad_id—method 1A would be the regression model based on average audience_size at the campaing_id level, and method 1B the regression that includes average audience_size at the campaign_group_id as explanatory variable

year = ad_id's year in ad_create_date

month = ad_id's month in ad_create_date

revenue = lifetime_legal_revenue

impressions = lifetime_legal_impressions

95. Third, I can continue building on the regression model above by including additional variables (terms to include in the regression) from all_ad_details and—as needed—ad targeting details, like country and age. The purpose of this is twofold: to exploit the information in average audience_size at the account level as well as improve the model's estimation accuracy. To the extent that there are ad_ids in all_ads_details that

do not appear in the ad targeting details data set and therefore have no targeting variables, I can still impute the missing Potential Reach values by including other available information from all_ads_details into the regression model.

96. The expected results from the methodologies outlined above are as follows. Of the records missing Potential Reach, I expect to be able to impute the Potential Reach values for about 21% after using the first two methods, and for about another 68% after using the third.

97. As a result, there would be a remaining 11% for which I will not be able to impute audience_size using the methodologies described above, because Facebook provided no audience_size information whatsoever for these records[31]. Instead, I will employ a fourth method: create a cohort of similar ad_ids based on relevant variables—other than an average audience_size—from ad targeting details and/or all_ad_details databases. The imputed Potential Reach value of an ad_id would be the average audience_size of the corresponding cohort. All variables would initially be used to create a set of contributing Potential Reach values. I would determine the set of relevant variables by assessing the set of targeting criteria that produces an average audience closer to observed Potential Reach values.

98. This fourth approach to imputing Potential Reach values is essentially a cluster analysis that would indicate which variables are important to clustering, as well as the clusters of ad sets that would serve as pools of ad sets that could then be used to impute missing values. This methodology is known as cold deck imputation[32], and has been used

---

[31] As a practical matter, it will not be necessary to compute the Potential Reach for 11% because beginning in January 2018, Facebook did not provide any Potential Reach numbers below 1000. *See* Section VII.C.

[32] Horton, Nicholas J.; Kleinman, Ken P. (2007-02-01). "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models". *The American Statistician*. 61 (1): 79–90, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839993/ ; *see also* Kalton, Graham (1986), "The treatment of missing survey data". *Survey Methodology.* 12:

by the U.S. Census Bureau to provide missing information about a group of households with specific characteristics through imputation to households that have the same characteristics but are not missing the key information.

99. In this way, using the four methodologies described above, I would be able to impute the missing Potential Reach numbers for the August 2014 – April 2019 time period. Similarly, contingent on Facebook providing ad targeting details, as well as all_ads_details, for the whole class period, I would be able to impute any missing Potential Reach values for the entire class. After I impute the missing Potential Reach numbers, I will be able to classify ads by range of Potential Reach, including identifying set of ads with Potential Reach below any threshold.

VII. OPINIONS: FACEBOOK'S POTENTIAL REACH IS INFLATED

100. In this section I apply the methodologies I discussed above. In Section VII.A I consider a fifth source of inflation to build on the formula in Section VI.D. I then use the US-only Potential Reach inflation rates in Appendix 2A to calculate a 44.5% rate of inflation for the initial default Potential Reach for all of US, that an advertiser would be shown on Facebook's Ads Manager. Those calculations are also in Appendix 2B.

101. In Section VII.B I calculate the probability that Potential Reach above a certain threshold is inflated by at least 10% by (1) running a Monte Carlo simulation using Facebook's rates for four sources of inflation which will confirm a normal distribution, and also (2) calculating the variance of the distribution, as well using a formula derived to express the variance of the inflation rate. Also in Section VII.B, I demonstrate that it is a statistical certainty that any Potential Reach of at least 1000 will be inflated by 10%, and

---

1–16, available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1986001/article/14404-eng.pdf?st=LSBGCOn4. Kalton, Graham; Kasprzyk, Daniel (1982). "Imputing for missing survey responses". *Proceedings of the Section on Survey Research Methods*. American Statistical Association, available at: https://www.istat.it/en/files/2014/05/1982-004-ASA.pdf

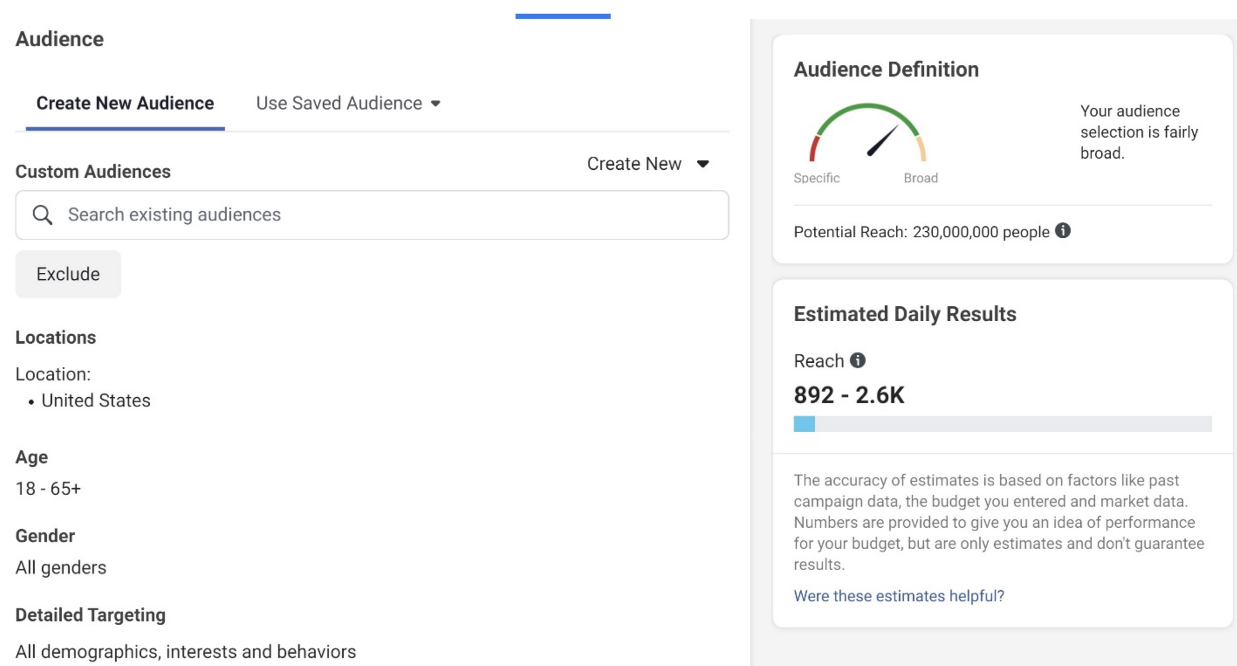
that the likelihood of inflation of at least 10% in a Potential Reach as low as 250 is well beyond any usual standard of testing.

102. Finally, in Section VII.C.1, I discuss how—from a practical perspective—it is not necessary for me to calculate the Potential Reach inflation of each individual advertiser, because based on Facebook's data 99% of the ads run between August 2014 and April 2019 had more than a 99.8% likelihood of Potential Reach inflation of at least 10%. Nonetheless, to illustrate the application of my methodology, in Section VII.C.2, I calculate the Potential Reach inflation using the named Plaintiffs' ad data sets.

A. Facebook's Initial Default Potential Reach for the United States Is Inflated

103. As noted in the Background section, at the outset, for advertisers who are located in the United States, Facebook's Ads Manager displays a default Potential Reach number for the entire United States. As seen in Figure 1 below, as of December 12, 2020, the default Potential Reach for the United States read "Potential Reach: 230,000,000 people."

Figure 1



104. The value shown as the default Potential Reach is what Facebook represents as the total number of people in the United States on Facebook. Using the methodology outlined in the previous sections, I have determined that this initial default Potential Reach number shown to all advertisers located in the United States is inflated.[33] As noted above, in addition to Fake accounts, Facebook's internal documents show that SUMA, Inactive, and Ineligible accounts also contribute to inflation of the Potential Reach any advertiser would be shown at the conclusion of demographic targeting, if any. However, the initial, default U.S. Potential Reach is impacted by one additional source of inflation: cross-platform inflation due to overlap between Facebook and Instagram accounts. This is because the default initial Potential Reach is displayed to an advertiser before the advertiser selects which Facebook platform (Facebook or Instagram) it wants to place the ad on. Because SUMA is a term referring to duplicate accounts on Facebook only (not Instagram), I will refer to these duplicates as "cross-platform" to avoid confusion.

105. Thus, for the purposes of calculating the inflation in the initial, default U.S. Potential Reach, I will use the following figures, obtained from Facebook's documents:

- **Fake** (5%) *See* SEC filings
- **Ineligible (US Only)** [redacted]
- **Inactive** (8.1%)[35]
- **SUMA (US Only)** [redacted]
- **Overlap with Instagram** [redacted] These are defined as instances where Facebook account holders also have Instagram accounts and where Facebook

---

[33] Facebook publicly acknowledges that its global user base (Monthly Active Users) is inflated. *See, e.g.*, 2018 10-K.
[34] *See* FB-SINGER-00170105
[35] *See* FB-SINGER-00258878, Facebook Q4 2019 Results, at https://s21.q4cdn.com/399680738/files/doc_financials/2019/q4/Q4-2019-Earnings-Presentation-_final.pdf

has not deduplicated these accounts when computing Potential Reach. To simplify my presentation, I refer to the Overlap with Instagram as **IGOverlap**.

106. Below are the steps to combine all five sources of default Potential Reach inflation.

1. The Overlap of Facebook and Instagram

107. From the Pew Research Study cited earlier, I can construct a table of the overlap between Facebook and Instagram users for the United States. Pew states that 68% of people in the U.S. use Facebook, that 35% of people use Instagram, that 47% of Facebook users have Instagram, and 91% of Instagram users have Facebook.[36]

108. Using these values, I put together the following table.

Table 5: Distribution of Accounts between Facebook and Instagram, US Population

U.S.	Instagram		
Facebook	Yes	No	Total
Yes	32%	36%	68%
No	3%	29%	32%
Total	35%	65%	100%

109. The margin values of 68% and 35% are cited directly, and a value between 31.7% 32.0% satisfies the conditions that 47% of Facebook users have Instagram and 91% of Instagram users have Facebook. I chose 32% in this range and the remaining values in the table are easily computed once this one value is determined.

110. This table has 29% of people who have no account with either Facebook or with Instagram. However, if I drop the 29%, I can compute the distributions of people with both accounts or one of the two accounts as shown in the next table, where the 71%

[36] Aaron Smith and Monica Anderson, "Social Media Use in 2018: Demographics and Statistics." Pew Research Center, March 1, 2018. Available at: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/.

(100% minus 29%) of the accounts are now the full population that would comprise the Potential Reach. Entries in this modified table sum to 100% of account holders.

Table 6: Distribution of Accounts, Facebook and Instagram, US Population, Conditional on Accounts That Exist

Modified	Instagram		
Facebook	Yes	No	Total
Yes	**45%**	51%	96%
No	4%		
Total	49%		

111. From the modified table, 96% of account holders would have a Facebook account, with or without Instagram and that there would only be an additional 4% of accounts that are only Instagram users with no Facebook. This number (4%) is needed in the next step of the calculations. I will refer to this total as IG~FB, so that I can include this value in my formulaic description of total accounts in Potential Reach.

2. Relating Accounts to People

112. In the second step, I relate accounts to people. Fake accounts relate to no people according to Facebook. SUMA and IGOverlap accounts relate to multiple counts associated with individual people. This can be summarized as a simple relationship that accounts for the origin of accounts but does not tally pieces by their contribution overall.

$$1)\ Potential\ Reach\ consists\ of\ Fake\ \&\ SUMA\ \&\ IGOverlap\ \&\ IGnoFB$$

113. I adopt the following assumptions related to the process of counting these accounts, namely:

iii. Fake accounts count as one each; they do not have duplicate Facebook (SUMA)accounts nor cross-platform accounts on Instagram (IGOverlap).

iv. The SUMA rate for each individual Facebook account is enumerated as 1, 2, 3, …; while most accounts will only have a single user, some have more than one. If no one had multiple accounts, SUMA rate would equal the number of people.

v. IGOverlap occurs only if there is a Facebook account and an Instagram account *and* the two accounts are not deduplicated in the Potential Reach computation.

vi. To be conservative, I assume there are no multiple Instagram accounts in IG~FB as there would be for SUMA because I do not have the rates of duplicate accounts on Instagram. [37]

vii. To be conservative, I assume there are no fake Instagram accounts in IG~FB.

114. Based on the above, I assume that IG~FB is only one account per person. This assumption does not apply to IGOverlap, since these are Facebook accounts and could contain SUMA accounts. With this information and these assumptions, I can rewrite equation 1, using Table 6, as follows:

$$2)\ Potential\ Reach - Fake$$
$$= (\%\mathrm{FB\&IG} + \%\mathrm{FB{\sim}IG}) * SUMA$$
$$+ (\%\mathrm{FB\&IG}) * IGOverlap$$
$$+ (\%\mathrm{IG{\sim}FB}) * IGnoFB$$

Where "~" connotes "not", and terms are defined as in Table 7 below.

Table 7: Terms for Inclusion in Facebook and Instagram

Terms	Instagram		
Facebook	Yes	No	Total
Yes	%FB&IG	%FB~IG	%FB
No	%IG~FB	%~FB~IG	%~FB
Total	%IG~FB	%~IG	100%

[37] Facebook documents indicate that multiple/duplicate accounts are actually even more common on Instagram than on Facebook. *See, e.g.,* FB-SINGER-00255393 at '96. However, in the absence of more robust data about Instagram-only accounts, I keep my assumptions conservative.

115. Actual accounts tied to people (Potential Reach minus Fake accounts) is equal to the sum of three components: 1) all SUMA accounts, 2) accounts on both Facebook and Instagram, where some portion overlap because they relate to the same person but are not deduplicated, and 3) accounts for people who are on Instagram, but not in Facebook.

3. Description of the SUMA Rate in Terms of People

116. The SUMA rate is the ratio of accounts to people for Facebook. This may occur in Instagram accounts, but there is no research provided by Facebook to document this, so I take a conservative approach by applying SUMA rate only to Facebook accounts. As already described in Section VI.D, the SUMA rate would be calculated as:

$$\begin{aligned} SUMA = [\, 1 * (\% \; of \; people \; with \; one \; account \; in \; FB) \\ + 2 * (\% \; of \; people \; with \; 2 \; accounts) \\ + 3 * (\% \; of \; people \; with \; 3 \; accounts) + \cdots] * People \end{aligned}$$

where the % of people with 1, 2, 3, … accounts sums to 100%.

117. A shorthand version of this calculation would be:

$$SUMA \; rate = People * [1 * p_1 + 2 * p_2 + 3 * p_3 + \cdots] = People * k$$

where p_j is the proportion of people with "j" accounts, and the sum of the p_j is 1.0, and "k" is the summation made by Facebook and available in its documents[38].

[38] FB-SINGER-00426295.

4. Description of IGOverlap in Terms of People

118. Facebook could match about [redacted] Instagram users to Facebook users.[39] But, the Pew study shows that 91% of Instagram users also have a Facebook account.[40] This means that [redacted] of Instagram accounts reported were duplicates of Facebook accounts. Per Table 6, in the United States about 45% of Facebook account holders have Instagram accounts. This means that [redacted] of U.S. accounts in Potential Reach are the same person with two accounts: one on Facebook and the other on Instagram.

5. Description of IG~FB in Terms of People

119. As discussed above, I assume that people who have Instagram accounts but do not have Facebook accounts have no Fake accounts (or they are tacitly included in Facebook's tally of Fake accounts) and no SUMA accounts. This means that IG~FB accounts are each a single person, with no SUMA inflation, but they can be Fake, Inactive, or Ineligible

6. Combining SUMA and Deduplicated IGOverlap Components for People

120. With these definitions, I rewrite equation 2 in terms of people as follows:

$$3)\ Potential\ Reach - Fake$$
$$= (\%\mathrm{FB\&IG} + \%\mathrm{FB{\sim}IG}) * \mathrm{SUMA} * People$$
$$+ (\%\mathrm{FB\&IG}) * \%Deduplicated\ Overlap * People$$
$$+ (\%\mathrm{IG{\sim}FB}) * People$$

[39] FB-SINGER-00125075.
[40] Aaron Smith and Monica Anderson, "Social Media Use in 2018: Demographics and Statistics." Pew Research Center, March 1, 2018. Available at: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/.

121. A variant of this equation would add "SUMA*People" in the last line in place of the term "People", to allow for SUMA on Instagram accounts. Not including this term makes this equation more conservative in estimating inflation.

7. Accounting for Ineligible and Inactive Accounts in Potential Reach

122. Both Inactive and Ineligible accounts need to be considered in determining Potential Reach inflation. Table 9 below shows a potential relationship between the two factors, since an account must be both Eligible and Active to be properly included in Potential Reach.

Table 8: Relationship between Eligible and Active

Global	Activity		
Eligibility	Active	Inactive	Total
Eligible	[redacted]	[redacted]	[redacted]
Ineligible			
Total	92%	8%	100%

123. I assume that the status of being an Active account and the status of being Eligible for an account are independent of one another, which allows me to estimate that, globally, [redacted] accounts are both Active and Eligible, meaning [redacted] of accounts are Inactive, Ineligible, or both.

124. This is a conservative assumption. If I assumed that an account could not be Inactive and Ineligible simultaneously, this would force the [redacted] in Table 4 to be zero. To maintain the marginal totals provided by Facebook, this would in turn force the proportion of active and eligible cases from [redacted] own to [redacted] and the values [redacted] and [redacted] to [redacted] and [redacted] espectively, increasing the inflation in my estimates.

8. Adding the Inactive and Ineligible to the Computation of Inflation

125. Equation 3 above gave the relationship between Potential Reach and People. Writing this again as equation 4 I get:

$$4)\ Accounts - Fake = [(\%\mathrm{FB}) * \mathrm{k}\ + (\%\mathrm{FB\&IG}) * \%DedupOverlap + (\%\mathrm{IG{\sim}FB})] * People$$

And solving for People, I get equation 5.

$$5)\ All\ People = \frac{Accounts * (1 - f)}{[(\%\mathrm{FB}) * \mathrm{k}\ + (\%\mathrm{FB\&IG}) * \%DedupOverlap + (\%\mathrm{IG{\sim}FB})]}$$

Where the value "f" is the rate reported by Facebook for Fake accounts.

126. However, the number of accounts must be reduced to only those that are Eligible and Active. Equation 6 is modified to account for only those people who may see an ad.

$$6)\ People\ Who\ May\ See\ Ad = \frac{Accounts * (1 - f) * (1 - g) * (1 - v)}{[(\%\mathrm{FB}) * \mathrm{k}\ + (\%\mathrm{FB\&IG}) * \%UndupOverlap + (\%\mathrm{IG{\sim}FB})]}$$

127. Here, "g" is the Ineligible rate, "v" is the Inactive rate, both provided by Facebook, and (1-g)*(1-v) is the proportion of Eligible and Active accounts, consistent with the computation presented in Table 8 above. Equation 6 is comprised of only values provided by Facebook, globally or by country.

128. Using equation 6, I can solve for Potential Reach inflation as the ratio of accounts to people who could potentially to see the ad. This is shown in equation 7:

$$7)\ Inflation = \frac{Potential\ Reach}{People\ Who\ May\ See\ Ad} = \frac{[(\%\mathrm{FB}) * \mathrm{k}\ + (\%\mathrm{FB\&IG}) * \%DedupOverlap + (\%\mathrm{IG{\sim}FB})]}{(1 - f) * (1 - g) * (1 - v)}$$

9. Computation of Inflation in Initial Default US Potential Reach

129. I use the following values to compute the inflation rate of the initial Default Potential Reach for the US:

Table 9: Values Needed to Compute Default US Potential Reach Inflation

Factor in Equation	Description	Source	Value
k	SUMA	FB-SINGER-00426295 analysis – Appendix 5	[redacted]
f	Fake Accounts	2018 10-K	5.0%
v	Inactive	FB-SINGER-00258878	8.08%
g	Ineligible	FB-SINGER-00170105	[redacted]
%Overlap	Overlap of FB & IG Not Corrected	FB-SINGER-00125075 & Pew study	[redacted]
%FB	%FB Accounts in Total Accounts	Table 6	96.0%
%FB&IG	%FB and IG Accounts (overlap)	Table 6	45.0%
%IG~FB	% Accounts Only from IG	Table 6	4.0%

130. Taking all five sources of inflation together, and assuming reasonable correlation and overlap[41] between the sources, I utilize the values above in equation 7:

Inflation = ([redacted]1.445

131. Thus, I conclude that Facebook's initial, default Potential Reach for the United States is inflated by at least 44.5%.

B. Potential Reach Is Inflated by at Least 10% at Convergence Thresholds

132. In this section I demonstrate that Facebook's Potential Reach is inflated by at least 10% at a convergence threshold. In particular, assuming a minimum Potential Reach

[41] A sensitivity analysis of the overlap between the five sources of inflation in the default US Potential Reach is in Appendix 1.

number at the conclusion of demographic targeting, I calculate to a reasonable degree of statistical certainty that the Potential Reach is inflated by at least 10%. The degree of statistical certainty increases as the minimum Potential Reach increases. At a threshold of 500, 1000 or 1100, the probability of receiving a Potential Reach inflated by at least 10% is greater than 99.99%. And even at a threshold of 250 Potential Reach, the probability of receiving at least 10% inflation is greater than 99.8%.[42]

133. To be clear, as I will demonstrate in Section VII.B2, my calculations of the Potential Reach inflation rates show that, based on Facebook's own data, even the *least* inflated demographic group has an inflation rate of over 19%—nearly double the illustrative 10% benchmark I use in my probability calculations. Thus, my demonstrative calculations of the probability of Potential Reach inflation are extremely conservative.

134. The approach set forth in this section is based on well accepted statistical principles. It is also intuitive. As noted in Section VI.H.2, according to the Central Limit Theorem, sums of random variables tend to the normal distribution as the size of the group being summed grows. An example of this statistical principle can be seen in epidemiological reports with respect to Covid-19. A recent report[43] noted that assuming just 20,000 persons were infected with Covid-19 in the United States (population 330 million), if a person attended an NFL football game with 75,000 people, the probability that someone at the stadium has Covid-19 is greater than 99%.

135. Here, as the size of relevant group– Potential Reach – increases, the inflation level is more likely to tend toward the baseline inflation. And because the baseline inflation on

---

[42] Facebook rounds Potential Reach numbers to the first two digits. *See* Expert Report of Dr. Atif Hashmi at ¶ 45; *see also* Rahul Bhandari Deposition, at 67:21–68:4; FB-SINGER-00084221, at '22. This means that Potential Reach numbers are sometimes further inflated due to rounding, or deflated due to rounding. However, the deflationary (and inflationary) effect of rounding can never be more than 5%. Therefore, rounding does not impact my analysis. *See* Appendix 14.

[43]COVID-19 Event Risk Assessment Planning Tool, available at https://covid19risk.biosci.gatech.edu

Facebook's platform is approximately 50%, the probability of reaching just 10% inflation is extremely high even at relatively small Potential Reach thresholds.

1. Monte Carlo Simulation Demonstrates Greater Than 10% Inflation

136. There is natural variation in Potential Reach, as Facebook only reports mean values for SUMA and Eligibility, but no indication as to how much numbers might vary within the groups for which they do report. As also discussed in Section VI.H, Potential Reach is properly expressed as the summation of indicator variables, that take on only two values: Fake\Genuine, Eligible\Ineligible, Active\Inactive; to be conservative, I will also treat SUMA as One Account\Two Accounts. These are also known as binomial variables. I can use properties of the binomial distribution and the Monte Carlo simulation methodology for accounts within a Potential Reach of a set size to determine the likely range of inflation within these groups, which I expect to be distributed as a normal variate. Using a Monte Carlo simulation run 10,000 times, I find the lower end of the range, where it is extremely unlikely that there are values of inflation below this lower end. Based on this, I can show that above a certain Potential Reach value all subgroups are inflated.

137. Since there are four sources of inflation, which can be represented as four binomials, the Monte Carlo simulation functions like flipping a coin four times, each time with a different chance of an outcome. In this case, I start with the probability that an account is Genuine, knowing that the rate of Fake accounts is 5%. The first coin flip comes up heads 95% of the time, tails 5% of the time. If tails, i.e. the account is Fake, we stop coin flipping since the other outcomes do not matter. If heads, i.e. the account is Genuine, then the next coin is flipped to determine whether the account is Eligible, with heads (the account is Eligible) occurring [redacted] the time. And so on with respect to Active and SUSA accounts.

138. This scenario of coin flips should be familiar. If I flip a fair coin 10 times, I can obtain a count of the number of heads that is anywhere from 0 to 10, with five heads the

most likely outcome, but with 0 or 10 heads as a possibility. In the same way, I can flip coins with specific probabilities of heads and tails and do so four times to develop the distribution of potential outcomes.

139. As an example for the purposes of this analysis, I run the Monte Carlo simulation for a set of 1100 accounts. Then I repeat this exercise 10,000 times to ensure a large enough sample of outcomes to have a clear picture of the distribution. This gives me 10,000 measures of inflation, all based on the same input probabilities as provided in Facebook's own documents. As a result, I obtain a distribution of outcomes, as shown in Chart 2A below. Chart 2B contains the same values as Chart 2A, but with the inflation value of 1.0 included on the x-axis.

Chart 2A: Distribution of a Sample Set of 10,000 Inflation Measures for Potential Reach of 1100

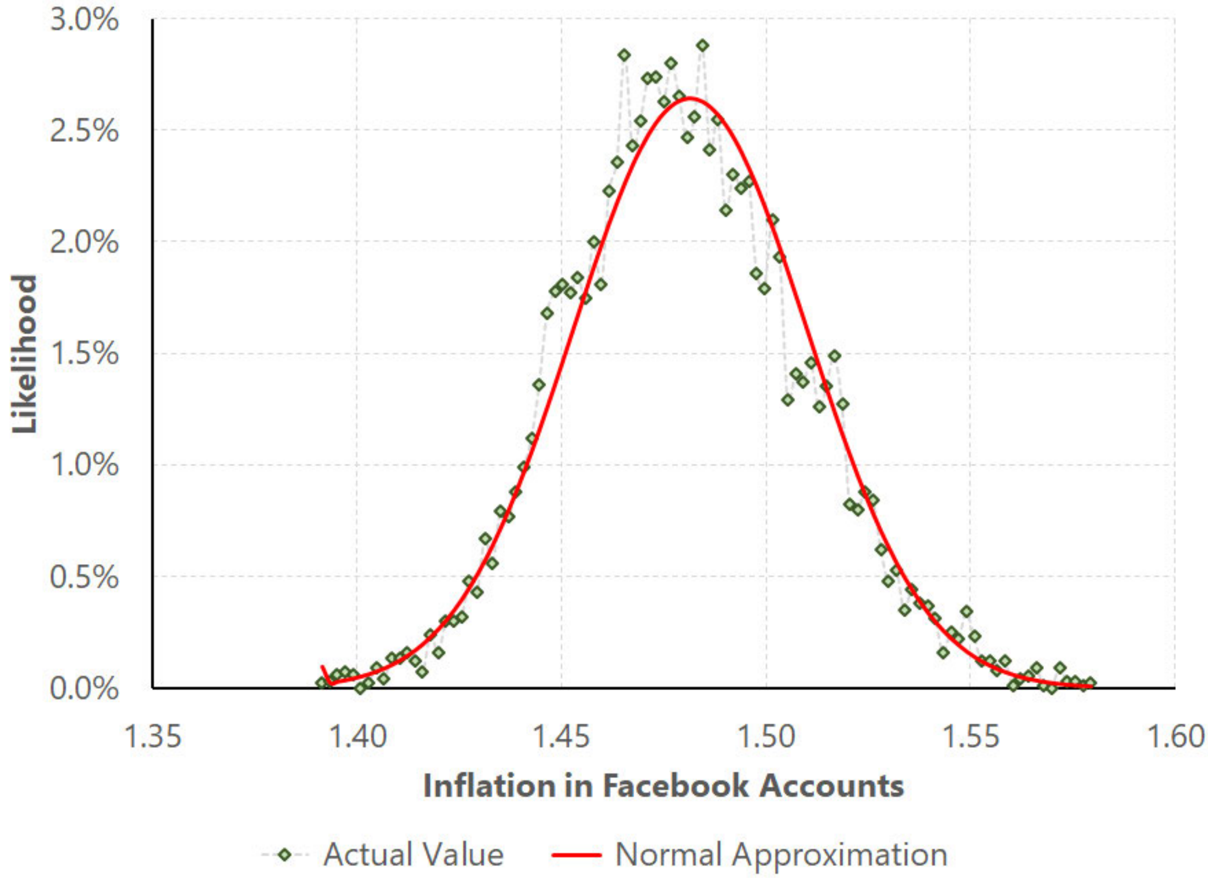
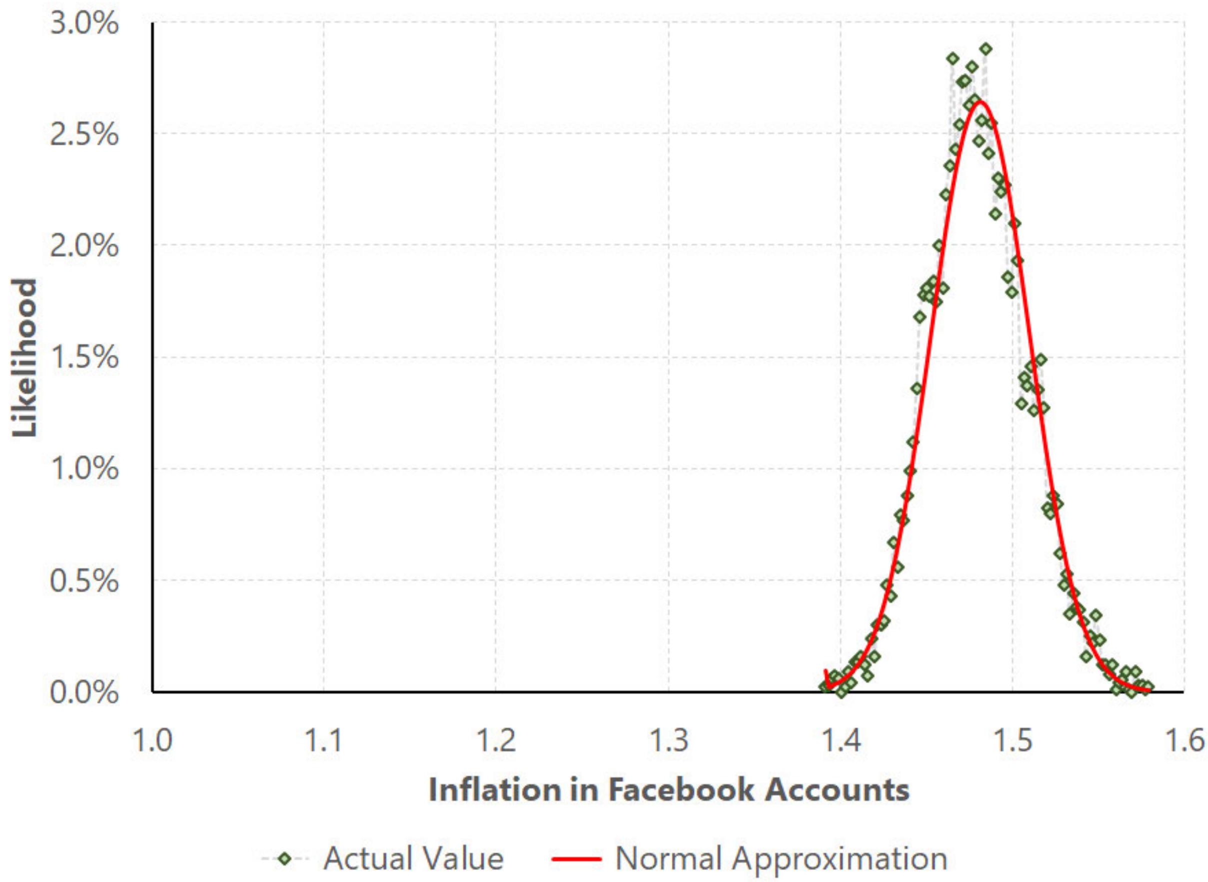


Chart 2B: Distribution of a Sample Set of 10,000 Inflation Measures for Potential Reach of 1100 –1.10 Inflation Rate Shown on X-Axis



140. Outcomes on Chart 2A (and 2B) show a clear pattern, with the center of the distribution at 1.47986, and a particular spread—the normal distribution, which I expected based on the normal theory arguments presented previously.[44] Here, the (visible) range of the distribution is from 1.39044289 to 1.580310881. This means that the center of the distribution is a 48% inflation, with the distribution ranging from 39% to 58% inflation at each end, respectively in the 10,000 outcomes. Notably, the inflation rate of below 10%

[44] In Appendix 12, I run an additional sensitivity analysis to confirm a normal distribution in this Monte Carlo simulation.

or 1.10 has such an infinitesimal likelihood of occurrence, that it is not even visible on either chart.

141. Thus, as Charts 2A and 2B show, given the rates of the four sources of inflation, the likelihood that a Potential Reach of 1100 is inflated by less than 10% is infinitesimal.

2. Probability Calculations Relying on Variance Calculations Also Show that the Likelihood that Potential Reach of 1100 is Inflated by Less than 1.10 Is Infinitesimal

142. Since, as discussed in Section VI.H, Potential Reach is a summation of the number of accounts that might be shown an advertisement, next, I compute the expected mean and variance of these Potential Reach summations (or, counts). I do so, because the mean and variance are the "sufficient statistics" that specify the shape and location of most probability distributions, including the normal distribution.

143. Based on Facebook's documents, I have the percentage of accounts that are Fake, Ineligible, Inactive, and SUMA. Because these percentages describe central tendencies for a random variable (i.e. inflation), I can use statistical methods to compute an expected value of inflation for any Potential Reach displayed by Facebook—to Plaintiffs or any other advertiser. Following the same statistical principles, I can also compute the variance and standard deviation associated with the inflation measure. With the variance, I can compute the likelihood that the inflation in a Potential Reach of a particular size is less than or equal to a fixed number.

144. If I write the equation for the inflation for a certain Potential Reach in terms of the sums of outcomes that might occur for each of the four factors, I get an estimate of inflation that is a random variable, T. I develop the variance equation in two steps by considering that inflation is a ratio with one random variable in the numerator and three

random variables in the denominator. The equation for a ratio of two random variables is available in multiple sources[45].

$$\mathrm{T}(x_s, x_f, x_v, x_g) = \frac{1 + \frac{x_s}{N}}{\left(1 - \frac{x_f}{N}\right)\left(1 - \frac{x_v}{N}\right)\left(1 - \frac{x_g}{N}\right)} = \frac{R}{W}$$

$$V(T) = \left(\frac{1+s}{(1-f)(1-v)(1-g)}\right)^2 \left[\frac{s(1-s)/N}{(1+s)^2} + \frac{Var(W)}{\{(1-f)(1-v)(1-g)\}^2}\right]$$

And

$$Var(W) = [(1-v)(1-g)]^2 \frac{f(1-f)}{N} + [(1-f)(1-g)]^2 \frac{v(1-v)}{N} + [(1-f)(1-v)]^2 \frac{g(1-g)}{N}$$

So

$$V(T) = \left(\frac{1+s}{(1-f)(1-v)(1-g)}\right)^2 \left[\frac{s(1-s)/N}{(1+s)^2} + \frac{f(1-f)/N}{(1-f)^2} + \frac{v(1-v)/N}{(1-v)^2} + \frac{g(1-g)/N}{(1-g)^2}\right]$$

$$V(T) = \frac{1}{N}\left(\frac{1+s}{(1-f)(1-v)(1-g)}\right)^2 \left[\frac{s(1-s)}{(1+s)^2} + \frac{f}{1-f} + \frac{v}{1-v} + \frac{g}{1-g}\right]$$

145. Thus, for a Potential Reach of a set size, N, I can write the inflation as it would be computed for any one Potential Reach if I knew the values for the number of SUMA, Fake, Inactive, and Ineligible Accounts in that one Potential Reach. Using the statistical

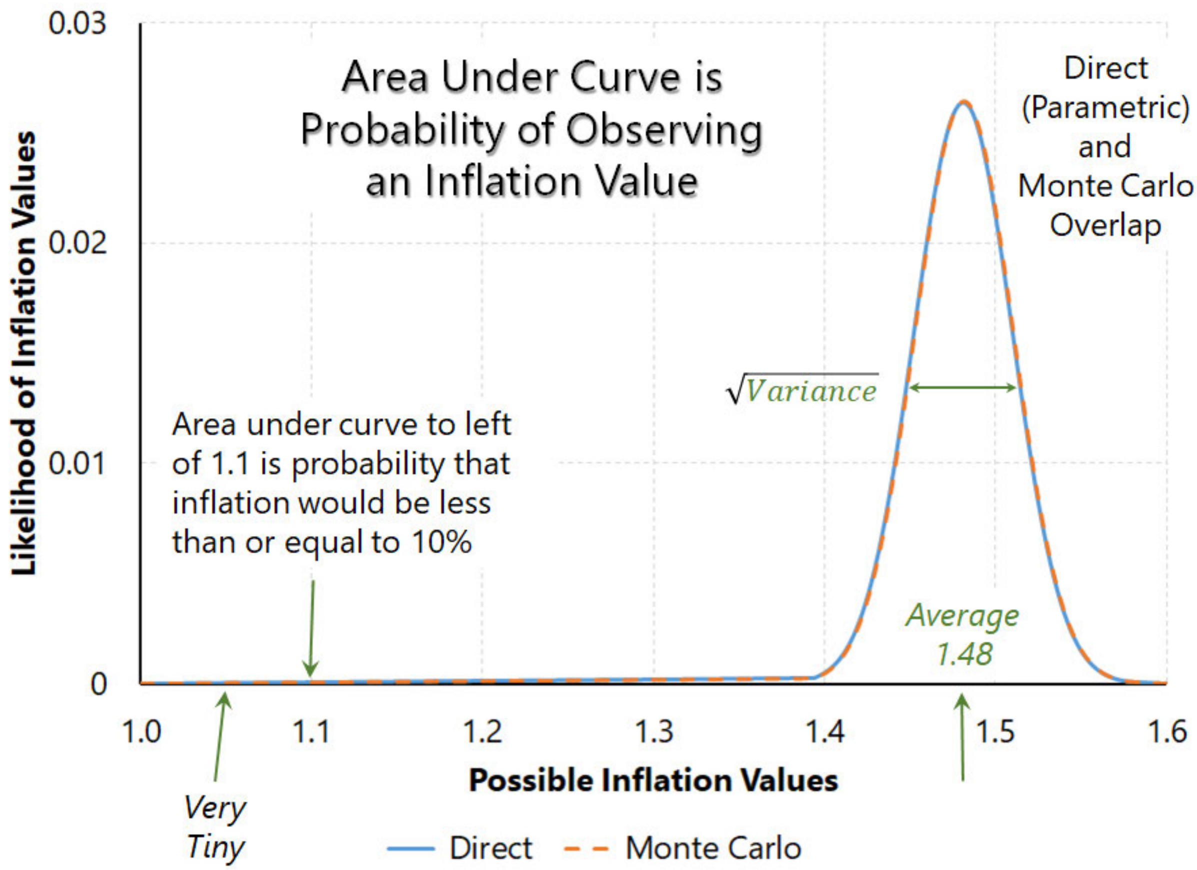
---

[45] For example, *Kendall's Advanced Theory of Statistics,* Wiley, London, 1998, 6th Edition, Volume 1, by Alan Stuart & Keith Ord

principles that follow from the information disclosed by Facebook, I can also compute the range of inflation that could be observed for a fixed Potential Reach. Here, as in Section VII.B.1, my computations use a Potential Reach of 1100 merely as an example.

146. In the final equation presented for V(T) above, the variability of inflation is a function of the size of the Potential Reach (the value N in this equation). Thus, as the Potential Reach increases, the range around the expected inflation decreases. This range is important for determining the likelihood that the true Inflation is greater than 10%. The following chart describes the factors that go into this computation.

Chart 3: The Bell Curve and Measuring Probabilities



147. In Chart 3, inflation is computed as 1.48, meaning that the Potential Reach is 48% greater than the number of people who could potentially see the ad to which the Potential Reach applies. The center of the bell curve represents the mean value computed for the

inflation. However, to ascertain the range of inflation, we also need to know the width of the bell curve. The width of the curve is expressed as the square root of the variance. The variance is the average distance of the curve from the midpoint line.

148. The area under the curve depicted in Chart 3 is the probability of various values of inflation, and the total area under the curve is 100% of the probability.[46] This means that the bell curve describes all possible values of inflation and the associated probabilities for a range of inflation values. In particular, one can see from Chart 3 that the probability of an inflation of less than 1.10 (10%) is far to the left in the chart and is extremely unlikely to ever occur – a probability of less than $3.5\text{x}10^{-40}$. The exponent of -40 means that this probability is less than 3.5 over a ten thousand quintillion quintillion.

149. For any value of inflation and the associated variance, I can compute a standardized statistic that allows me to compute the likelihood through reference to the values of the standard normal distribution. Computing the mean and variance as described in the preceding paragraphs, I can compute a standardized "z" value as:

$$z = \frac{1.1 - Mean\ Inflation}{\sqrt{Variance\ Inflation}}$$

150. Since the mean inflation values are always greater than 1.1 (using 10% inflation threshold and the values for the different sources of inflation obtained through discovery), the value of z is always negative. The more negative z is, the less likely it is that there would be a value of inflation that is less than 1.1. Also, as the Potential Reach increases, the value of z increases meaning that for a fixed inflation value and increasing N, it is more and more unlikely that inflation would ever be below 1.1.

[46] There are several methods used to calculate the area under the normal curve. Regardless of the method chosen, they all attempt to do exactly the same thing, although through different approaches. As such, the probabilities given above might differ from probabilities computed using a different method, but the differences must be negligible.

151. Because z is distributed as a normal variate with mean zero and variance 1.0, I can utilize 1.10 for this formula in the numerator (regardless of the mean and variance) and this gives me the test statistic z. I can now look up the value z on the standard normal distribution and find out how much probability is below this value.

152. For instance, I can calculate the likelihood that the five demographic groups on Facebook (in countries where there is a Potential Reach of at least 1100) with the lowest rates of Potential Reach inflation are not inflated by at least 10%. As discussed in Section VII.F, I calculated the Potential Reach inflation rates by country and age groups in Appendix 2A. Taking the five demographic groups that have the lowest rates of Potential Reach inflation and utilizing the mean predicted inflation into the variance equation, I get the following probabilities in Table 10. [47]

Table 10: Likelihood of Inflation Less Than 1.1 In 5 Demographics with Lowest Rates of Potential Reach Inflation

Country	age group	Predicted Inflation	Variance of Inflation at n=1100	Prob Inflation 1.1 or Less
Norway	45-49	1.194	0.0002	4.72E-10
Norway	50-54	1.212	0.0003	2.65E-12
Norway	40-44	1.219	0.0003	2.58E-13
Norway	55-59	1.223	0.0003	1.02E-13
Denmark	50-54	1.224	0.0003	5.99E-14

153. Next, of these five I can take the demographic with the lowest Potential Reach inflation and calculate the likelihood that inflation is not at least 10%, using a range of Potential Reach thresholds. Again, it is worth noting that my calculations in Table 10 show that, even the *least* inflated demographic group has an inflation rate of over 19%—nearly double the illustrative 10% benchmark I use in my probability calculations in Table 11

[47] Details of calculating the values in Tables 10 and 11 are in Appendix 6.

below. For some perspective, according to Facebook's data, of all the ads that ran in 2018, only 61 were targeted at this least inflated demographic of Norway, ages 45-49; of those only 48 had a Potential Reach value, which ranged from 1000[48] to 86,482. For that reason, my calculations of the probability of Potential Reach inflation in Table 11 are extremely conservative.

Table 11: Likelihood of Inflation Less Than 1.1 In Demographic with Lowest Potential Reach Inflation

Country	Age Group	Potential Reach	Predicted Inflation	Variance of Inflation	Probability of At Least 1.1 Inflation	Probability of Inflation 1.1 or Less
Norway	45-49	250	1.194	0.00103	>99.82%	1.77E-03 (1 in 566)
Norway	45-49	500	1.194	0.00051	> 99.99%	1.85E-05 (1 in 54 thousand)
Norway	45-49	1000	1.194	0.00026	> 99.99%	2.71E-09 (1 in 3.7 hundred million)
Norway	45-49	1100	1.194	0.00023	> 99.99%	4.72E-10 (1 in 2 billion)
Norway	45-49	1500	1.194	0.00017	> 99.99%	4.50E-13 (1 in 2 trillion)
Norway	45-49	2000	1.194	0.00013	> 99.99%	7.90E-17 (1 in 12 quadrillion)
Norway	45-49	2500	1.194	0.00010	> 99.99%	1.43E-20 (1 in 70 quintillion)

154. As shown in Table 11 above, the likelihood that even a very small Potential Reach of 250, in the lowest inflated demographic, is inflated by less than 10% is more than 99.8%. At 500, that likelihood is one in 54,000. And as soon as Potential Reach is set at 1000, that likelihood jumps to one in 370,000,000. And, because the lowest Potential Reach value

[48] Again, because ads that were created after December 31, 2017 displayed a Potential Reach of no less than 1000 on Ads Manager.

provided after January 2018 is 1000, the likelihood that any ad set received a Potential Reach inflated by less than 10% is no greater than one in 370,000,000.[49]

155. Given the above probabilities at the county and age level, it is a statistical certainty that regardless of more granular targeting criteria any Potential Reach of at least 1100, 1000, or 500 will be inflated by at least 10%. Even at a Potential Reach of 250, the likelihood of inflation of at least 10% is well beyond any usual standard of testing.

C. Applying the Methodology to Plaintiffs and Each Class Member

1. As a Practical Matter, It Not Necessary to Calculate the Potential Reach Inflation for Each Advertiser

156. I can apply the methodology described above to any advertiser. Using the inflation formula in Section VI.D, I can calculate the inflation of any advertiser's Potential Reach. And as discussed in Section VI.I, to the extent Potential Reach numbers are missing, I can determine them as well. Finally, as discussed in Section VII.B, I can determine the probability that the advertiser's Potential Reach is inflated by at least 10%.

157. However, as shown in Table 11, in even the lowest inflated demographic (Norway, ages 45-49) the likelihood of inflation of at least 10% in a Potential Reach as low as 250 is well beyond any usual standard of testing (i.e. a 5% or 1% level of significance). According to Facebook's data, [50] of the ads that ran between August 2014 and April 2019, only 4.2% had a Potential Reach of below 1000; only 2.9% had a Potential Reach of below 500; and only 1.9% had a Potential Reach of below 250.

---

[49] For perspective, the chance of being struck by lightning is approximately 1 in 500,000. *See, e.g.,* "Natural Disasters and Severe Weather – Lightning." Centers for Disease Control and Prevention, available at: https://www.cdc.gov/disasters/lightning/victimdata.html.

[50] Facebook produced information about advertisements on Facebook, August 2014 – April 2019, as a table called all_ads_details, FB-SINGER026. The percentages here are calculated using only the ad_ids for which Facebook actually provided Potential Reach values and, again, is appropriately limited only to ads that actually ran (i.e., had a non-zero/positive revenue values).

158. This means that assuming even the lowest inflation rate Facebook reported at the country and age level, 98.1% of ads run between August 2014 and April 2019, had a Potential Reach of 250 or above, and thus, at minimum, had a 99.82% likelihood of at least a 10% inflation in Potential Reach. This probability is sufficient to conclude that 10% inflation of Potential Reach is a near statistical certainty. I could achieve an even higher probability, of more than 99.99%, by setting the threshold at Potential Reach of 500. Again, assuming even the lowest inflation rate Facebook reported at the country and age level, Facebook's own data shows that 97.1% of ads run between August 2014 and April 2019 had a Potential Reach of 500 or above, and thus, per Table 11, had a more than a 99.99% likelihood of at least a 10% inflation in Potential Reach for each advertiser.

159. But the probabilities above do not yet account for the fact that, regardless of what Facebook calculated as the Potential Reach on the backend, any ad created on or after January 1, 2018 would, at minimum, display a Potential Reach of 1000 on Ads Manager. As I understand from Dr. Hashmi's expert analysis of Facebook's source code, that after December 31, 2017, the Potential Reach displayed on the Ads Manager interface was set to a minimum threshold of 1000, regardless of the Potential Reach value that was calculated by Facebook on the backend.[51] [52] Thus, the only ads that could be subject to a lower than 99.82% probability of 10% Potential Reach inflation are ads that were created before January 1, 2018 (or August 2014 – December 31, 2017) and had a Potential Reach of less than 250.

160. Based on Potential Reach records produced in FB-SINGER-026, of the ads that ran between August 2014 and April 2019: only 2.2% were created before January 1, 2018

[51] Expert Report of Dr. Atif Hashmi, at ¶¶ 43-45.

[52] I also understand that ads that used Custom Audiences in that time period may have had Potential Reach minimums even higher than 1,000 for privacy-related reasons *See id.; see also* Rahul Bhandari Deposition at 66:2-67:20.

and had a Potential Reach below 1000, only 1.5% were created before January 1, 2018 and had a Potential Reach below 500; and only 1% were created before January 1, 2018 and had a Potential Reach below 250.

161. Thus, although I can calculate the Potential Reach inflation for any advertiser (and demonstrate so in Section VII.B.2), as a practical matter, it is not necessary to calculate the Potential Reach inflation for each individual advertiser. At a 250 Potential Reach threshold, 99% of the ads run between August 2014 and April 2019 each had at least a 99.82% likelihood of Potential Reach inflation of at least 10%. Thus, for 99% of ads the likelihood that Potential Reach will be inflated by at least 10% is well beyond any usual standard of testing (i.e. a 5% or 1% level of significance). The same is true if I extrapolate this to present day, assuming (extremely conservatively) that the percentage of pre-2018 ads with a Potential Reach of 250 remains 1%, even though no one received a Potential Reach number below 1000 after December 31, 2017.

2. <u>Nonetheless, I Can Calculate Potential Reach Inflation for Any Advertiser</u>

162. As an example and to illustrate my methodology, I will calculate the Potential Reach inflation for the Plaintiffs.[53]

163. Facebook produced Potential Reach and other transactional-level data for Plaintiffs in all_ads_details data sets: FB-SINGER-00180465 (DZ Reserve) and FB-SINGER-00180468 (Cain Maxwell). Facebook also produced the ad targeting criteria data for Plaintiffs: FB-SINGER-00180469 (DZ Reserve) and FB-SINGER-00180472 (Cain Maxwell).

164. First, I used the data produced by Facebook to impute any missing Potential Reach values for each Plaintiff. Then I applied the Potential Reach inflation rates by

[53] To further demonstrate the applicability of my methodology, I apply the analysis to three random ad sets from 2018. See Appendix 15.

country and age group as calculated in Appendix 2A to the targeting data produced for the Plaintiffs.

165. Plaintiffs' all_ads_details data sets, FB-SINGER-00180465 (DZ Reserve) and FB-SINGER-00180468 (Maxwell), include daily records of each Plaintiff's ad_id, so that ad_id-date (daily record) is the most granular unit of analysis. Both audience_size and lifetime_legal_revenue—among others—are provided for each ad_id-date.

166. First, I imputed any missing Potential Reach values at the ad set (campaign_id) level. To do so, I aggregated Potential Reach values at the ad_id-date level to obtain the Potential Reach at the ad_id level. Then, I aggregated Potential Reach at the ad_id level to calculate the values at the campaign_id level. Within ad_id, the correlation of Potential Reach among ad_id-dates was 0.99. Similarly, within campaign_id, the correlation of Potential Reach among ad_ids was also 0.99.

167. Given the available data, I was able to impute the missing Potential Reach values by aggregating the ad_ids and using the average as the Potential Reach value at the ad set level. To ensure robustness of data when aggregating Potential Reach averages, I also included Potential Reach values for ads that were never run (and thus did not have a positive lifetime_legal_revenue associated with them).

168. Next, I calculated the Potential Reach inflation for each Plaintiff at the ad set level. The ad targeting data for each Plaintiff, FB-SINGER-00180469 (DZ Reserve) and FB-SINGER-00180472 (Cain Maxwell), provides the country and age group targeted at the ad_id level. For each ad_id, I matched the targeting criteria to the corresponding Potential Reach inflation rates as calculated in Appendix 2A. I then calculated the average Potential Reach inflation rate and imputed it as the inflation for that ad_id. To calculate Potential Reach at the ad set (campaign_id) level, I calculated the average Potential Reach inflation of the ad_ids in a given ad set.

169. My calculations of Potential Reach and Potential Reach inflation rates for Cain Maxwell at the ad set level are in Appendix 8. Maxwell's Potential Reach ranged from

2,433,637 to 49,773,433. The Potential Reach inflation for Maxwell ranged from 1.3194 to 1.3270, or 32%-33%. And, based on my calculations above, I can state to a statistical certainty that every Potential Reach number received by Maxwell was inflated by at least 10%. The probability that Maxwell did not receive at least 10% inflation is less than 1 in 70 quintillion.

170. My calculations of Potential Reach and Potential Reach inflation rates for DZ Reserve at the ad set level are in Appendix 9.[54] DZ Reserve's Potential Reach ranged from 1000[55] to 2,147,483,647 (which would have been rounded to the first two digits). Potential Reach inflation for DZ Reserve ranged from 1.2796 to 1.4799 or 28%-48%. And, based on my calculations above, I can state to a statistical certainty that every Potential Reach number received by DZ Reserve was inflated by at least 10%. The probability that DZ Reserve did not receive at least 10% inflation is no less than 1 in 370,000,000.

---

[54] For DZ Reserve, I allocated 23,920 ad_ids (ads) into 17,878 campaign_ids (ad sets), out of which 16,103 had positive revenue. I estimated any missing Potential Reach values using the average aggregation method described above. Most of DZ Reserve's ads (about 86%) targeted all countries and all ages, and another 13% of them targeted either all countries or all ages. So, despite the relatively large number of ad_ids and campaign_ids, there was not much variability in targeting types 1-country and 2-age.

[55] FB-SINGER-00180465 shows that for DZ Reserve, out of 17,878 ad sets, 87 ad sets had a Potential Reach of below 1100, with 60 of them being below 500. However, the all_ads_details table, including FB-SINGER-00180465, has the Potential Reach calculated by Facebook, not necessarily what is displayed on Ads Manager. As I understand from Dr. Hashmi's expert report, between January 1, 2018 and through the April 6, 2019, the Potential Reach displayed on the Ads Manager interface was set to a minimum threshold of 1000, regardless of what the Potential Reach value that was calculated by Facebook on the backend. Expert Report of Dr. Atif Hashmi, at ¶¶ 43-45. I also understand that ads that used Custom Audiences in that time period may have had minimums even higher than 1,000 for privacy-related reasons. *See id.; see also* Rahul Bhandari Deposition at 66:2-67:20. Since DZ Reserve bought ads after January 1, 2018, the minimum Potential Reach number that would have been displayed on Ads Manager would have been 1000. In addition, as discussed in Appendix 14, any Potential Reach calculated on the backend would have been rounded to the first two digits when displayed on Ads Manager.

171. This is consistent with Section VII.B, in which I determined that it is a statistical certainty that any given Potential Reach above 1000 or 500 is inflated by at least 10%, and that even at a Potential Reach of 250 the likelihood of a 10% inflation is well beyond the usual standards of testing.

172. The calculations above demonstrate the application of my methodology. Just as I was able to calculate the Potential Reach inflation for the Plaintiffs, I can use the same methodology (if appropriate, combined with additional methodologies described in Section VI.I) to calculate the Potential Reach inflation for any other advertiser.

VIII. FACEBOOK'S OTHER INFLATION-RELATED EFFORTS DO NOT IMPACT MY MODEL

173. In this section, I briefly address two miscellaneous considerations pertaining to Facebook's calculations of Potential Reach and why they do not affect my model.

A. Facebook's Age Modeling Does Not Ameliorate Potential Reach Inflation and Thus Has No Impact on My Model

174. In FB-SINGER-00426295, the same spreadsheet in which Facebook produced the effects of SUMA inflation on a set of Potential Reach values for 11 age groups in 240 countries, Facebook also produced the effects of age modeling on Potential Reach inflation.[56] I used the age-adjusted and SUMA-adjusted Potential Reach values Facebook produced and re-ran the inflation values. But, I did not include the effects of Facebook's age-modeling in calculating the Potential Reach inflation for the following reasons.

175. First, I understand that Facebook decided that it could not implement age modeling [redacted].[57] Assuming that Facebook was correct in this

---

[56] *See, e.g.,* FB-SINGER-00080023 at slides 7-12.

[57] FB-SINGER-00264321

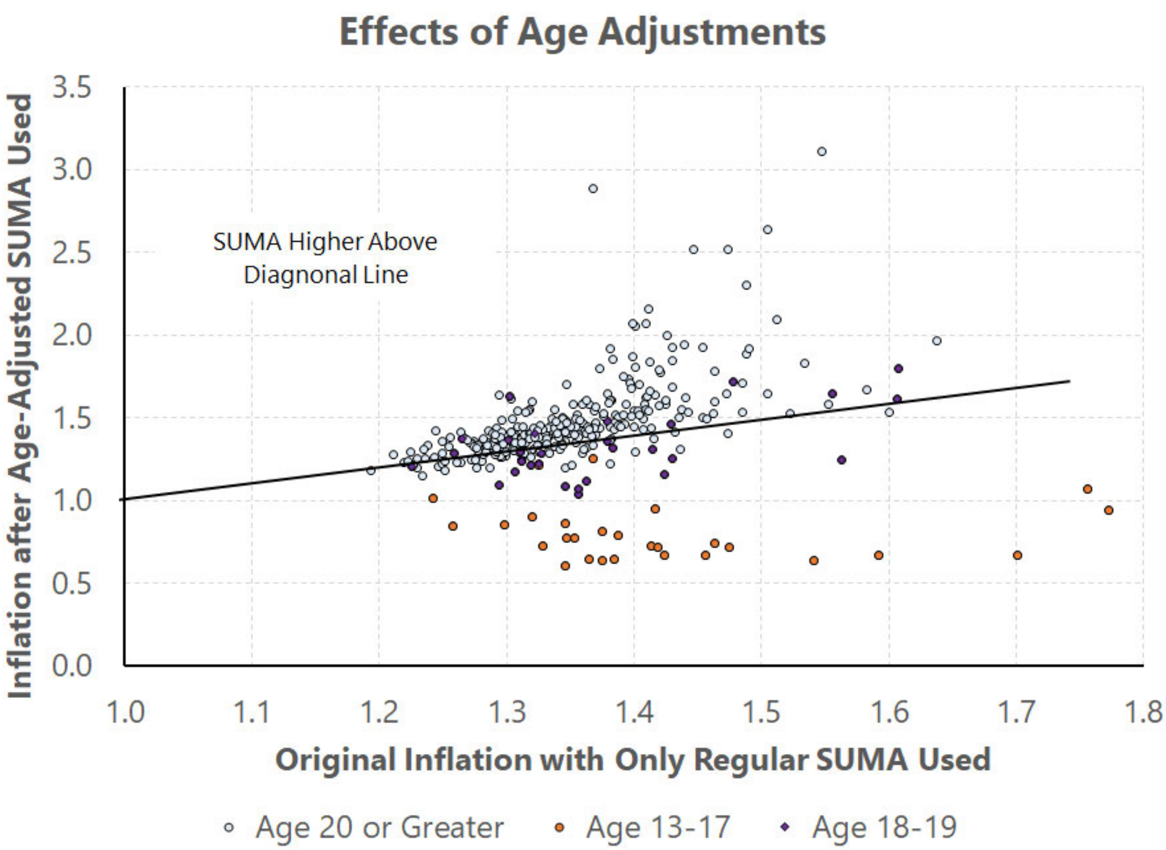
analysis, it would not make sense for me to rely on a model that could not be legally implemented, and indeed was not.

176. Second, even if I were to implement age modeling in my inflation model, Facebook's approach (as reflected in FB-SINGER-00426295) is not proper. As discussed below, Facebook's age modeling would have simply moved users from one age bracket to another. Instead of merely redistributing users, Facebook should have used age modeling to remove from age brackets the users whom Facebook identified as not belonging in those age bracket.

177. Chart 4 displays the original Potential Reach inflation numbers computed with only the SUMA adjustment on the horizontal axis versus the inflation numbers with SUMA and age adjustments and marked for the age groups 13-17 and 18-19.

[see next page for Chart 4]

Chart 4: Distribution of Inflation Before and After Age-Adjustments[58]



178. As seen in Chart 4, for age group 13-17[59], the Potential Reach inflation is lower in all countries. For age group 18-19, the Potential Reach inflation is lower for half the countries and higher for the other half the countries. However, for all other age groups, 20+, Potential Reach inflation *increases*.

---

[58] In Chart 4, two values for ages 13-17 are excluded because the original inflation was much higher than 1.8 but the new inflation value was lower than 1.5. This attenuated the chart, making it difficult to see the results for the other 328 observations that are displayed.

[59] Notably, ads targeted exclusively at the 13-17 age group represent a very small fraction of the class. For instance, of all the ads that ran in 2018 only 0.1% of the ads were targeted at this age group. Facebook's attempt to correct inflation of this age group, even if it had been successful, would have a minimum impact on inflation across the board.

179. This result comes about because the age adjustments are zero-sum, meaning that if accounts are moved out of the 13-17 age group, they are redistributed to other age groups in the same country. This means that when an advertiser targets a country but does not specify any age groups, there is no difference in Potential Reach due to the age-adjustments: the numbers are redistributed among the age groups but have no overall effect on the level of Potential Reach inflation due to SUMA within a country. As a result, if Potential Reach inflation goes down in one age group, like 13-17, the inflation must go up in one or more other age groups.

180. After accounting for age groups 13-17 and 18-19, there are 57 out of 330 other values where inflation declined (e.g. Australia age 30-34, or Vietnam age 55-59), though by much smaller amounts. The majority of values increased.

181. Based on the above analysis, I conclude that Facebook's age modeling did not decrease Potential Reach inflation but simply redistributed it among other age groups in the same country. Therefore, I did not consider the age modeling in my model and it has no impact on its applicability. If asked, I could compute the inflation rates using Facebook's age modeling, but I do not believe it is appropriate to do so for the reasons stated above

B. Facebook Had No Material Modifications to Potential Reach Calculations Since August 2014, and the Modifications It Made Have No Impact on My Model

182. Another reason why my model can be applied from August 2014 to the present is that Facebook has made few material modifications to its calculations since August 2014.

183. In March 2019, Facebook switched from calculating Potential Reach based on ad requests to calculating it based on ad impressions.[60] This was in part a change in what

---

[60] Facebook's Response to Interrogatory No. 3; Rahul Bhandari Deposition at 37:8-38:8.

made an account Eligible to view ads (and therefore included in Potential Reach).[61] This change is already part of my sensitivity analysis in Appendix 11, which shows that even if the rate of Ineligible accounts was down to 0%, the Potential Reach would still be very inflated.

184. Other than the March 2019 change, Facebook only made two other modifications that it believes were material.[62] Having reviewed the other two changes made to Facebook's calculation of the Potential Reach, I do not believe that either would be appropriate to consider in my model.

185. First, in December 2016 Facebook switched from a platform called Inventory Estimation ("InvEst") to a platform called "Cocoon" to calculate Potential Reach.[63] However, according to Facebook's documents, Cocoon's estimates are only much more accurate than InVest for Potential Reach under 1,000.[64] As discussed in Section VII.C.1 above, based on the August 2014 – April 2019 time period alone, very few ads had a Potential Reach of below 1000: only [redacted] Moreover, starting in 2018 Facebook rounded any Potential Reach number below 1000 to 1000, so only [redacted] of ads run could even have shown a Potential Reach below 1000—with an even smaller percentage (those ads that were run December 2016 – December 31, 2017) affected by the switch from InvEst to Cocoon. Finally, my opinion regarding Potential Reach inflation ultimately sets a Potential Reach threshold of 1000 and above—though as I show in Sections VII.B.2 and VII.C.1, it can certainly be set to a much lower threshold. Therefore, any differences in Potential Reach calculations made to a tiny fraction of all the ads that were run August 2014 – April

---

[61] *See* FB-SINGER-00184892.

[62] Facebook's Response to Interrogatory No. 3; Rahul Bhandari Deposition at 37:8-38:8.

[63] *Id.*

[64] FB-SINGER-00183123.

2019—and which are an even tinier fraction of the entire class period which I understand to be August 2014 through the present—are not relevant to my conclusions.

186. Because the impact of this December 2016 change would have been virtually nonexistent, it was not appropriate to consider in my model.

187. Second, in October 2018 Facebook switched from a heuristic model to [redacted]. However, Facebook's own documents state that [redacted].[66] And, as discussed in Sections V and VI.B, Facebook itself has taken the position that [redacted]. Furthermore, the switch to [redacted][67], which is too small given the levels of inflation found in my analysis. As discussed in Section VII.B, while I can calculate the inflation of each advertiser, even without conducting those calculations, I observe that 99% of advertisers between August 2014 and April 2019 alone have a 99.82% probability of having received a Potential Reach inflated by at least 10%—and likely the inflation is much higher since my calculations are very conservative and use the least inflated demographic (Norway, ages 45-49), which itself had a nearly-twice-as-large inflation of over 19%. In addition, according to Facebook[68], [redacted] of its revenue comes from ads that are targeted at the country level and [redacted] of revenue comes from ads that are targeted at the country or region level, with only [redacted] coming from targeting as granular as city-level. According to the targeting data produced for 2018 in FB-SINGER-035, 62.7% of the ads that ran were targeted on the

[65] Facebook's Response to Interrogatory No. 3.
[66] FB-SINGER-00278872; Rahul Bhandari Deposition at 53:8-20.
[67] FB-SINGER-00278872, at '85.
[68] FB-SINGER-00278872 at'76.

region level or higher. If necessary, I can simply segregate out the ads that were targeted at a more granular level for the entire class; my inflation analysis, however, would not change.

188. Because the impact of this October 2018 change would have improved a level of targeting Facebook itself considered unreliable, and the impact been so small compared to the levels of inflation I calculated, this change was not appropriate to consider in my model.

IX. CONCLUSION

Based on the foregoing analysis, I conclude that:

1) Facebook's initial, default Potential Reach number for the United States is inflated by at least 44.5%.
2) Every final Potential Reach at or above 1000 provided by Facebook to advertisers is inflated by at least 10%.
3) Plaintiff Maxwell's Potential Reach was inflated by 32%-33%, and the chance that he did not receive at least 10% inflation is less than 1 in 70 quintillion. Plaintiff DZ Reserve's Potential Reach was inflated by 28%-48%, and the chance that he did not receive at least 10% inflation is, at a minimum, 1 in 370,000,000.
4) The methodology outlined herein is the same whether applied to the named Plaintiffs or any advertiser, and can be used to determine the inflation rate for a Potential Reach at any threshold.

December 22, 2020

Charles D Cowan

Charles D. Cowan, Ph.D.

X. Exhibit 1: Resume

Charles D. Cowan is Managing Partner of ANALYTIC FOCUS LLC. Dr. Cowan has 40 years of experience in statistical research and design. He consults for numerous public and private sector entities on the design, implementation, and evaluation of research and the synthesis of statistical and sampling techniques for measurement.

Dr. Cowan has designed some of the largest and most complex research programs conducted by the Federal Government, including the Post Enumeration Program conducted by the Bureau of the Census to evaluate the 1980 Decennial Census, the Economic Cash Recovery valuations conducted by the Resolution Trust Corporation in 1990-95, and many evaluation studies conducted for the Justice Department, the Department of Defense, the Department of Housing and Urban Development, and the Treasury Department. He has provided expert advice to corporations and government agencies on the incorporation of complex research designs in demographic and economic measurement problems, including:

- Development of procedures used by the Resolution Trust Corporation and the FDIC for determination of the value of all assets held by the RTC\FDIC taken from failed banks and S&Ls. Results from this research were used in quarterly reports to Congress on the loss to the American taxpayer that resulted from these failures. These estimates of anticipated recoveries on assets were also used by the RTC and FDIC for financial reporting, leading these agencies to their first clean opinions from the GAO in their annual review of agency financial statements.
- Establishment of audit and sampling methods to determine the completeness and reliability of reporting and record systems. These procedures were used to both expand and streamline bank examinations for safety and soundness and also compliance measurement for the FDIC. These sampling techniques are applied in the audit of Federal agencies concerned with regulatory review of operations and systems, and related systems for banks, regulatory agencies, and law firms;
- Application of econometric and biometric procedures for measurement of credit risk in large portfolios of loans. These models are frequently used for a variety of purposes within financial institutions, such as the pricing of loans, the management of customers long term, decision making on workouts for delinquent loans, and for establishment of economic and regulatory reserves.
- Evaluation of research conducted for the Department of Defense, for the National Institutes of Health, and for the Department of Agriculture, each in response to

Congressional inquiries on the validity of published results, and also for defendants in lawsuits involving evidence proffered by plaintiffs in furtherance of their suit.

- Model fitting and development of projection methods to measure the likelihood of loss or errors in recording in loans held by banks or put up for auction; measurement of the likelihood of fraud and/or noncompliance in systems, including bank holding companies, trading activities for brokers, and systems for compliance with health department and judicial requirements;
- Incorporation of population demographic models with financial assessment models to predict risk for insurance companies and corporations in terms of number and value of potential claims in mass tort litigation.
- Development of procedures used by the Bureau of the Census for apportionment of population for revenue sharing purposes and the estimation of the undercount in the Decennial Census of Population and Housing. These procedures include application of capture-recapture methods to measure the size of the undercount in the decennial census, use of network sampling as an alternative measure for population size, and measurement of the reliability of data collected in the Census.
- Development of statistical methods to quantify the size of populations, including nomadic populations for the Census of Somalia, the undercount and overcount in the Census of Egypt, the number of missing children in Chicago, IL, and the number of homeless persons and families needing services in several large cities with transient populations.

Dr. Cowan teaches graduate and undergraduate courses in survey methods, statistics, and computer methods for analysis. He is the co-author of two books, one on evaluation of survey and census methods and one on econometric measures related to the welfare of the U.S. economy. He has written numerous articles on statistical methods, sampling, rare and elusive population research, and optimization techniques.

Prior to cofounding ANALYTIC FOCUS LLC, Dr. Cowan was a Director with ARPC and with Price Waterhouse, where he specialized in financial research and audit sampling. From 1991 to 1996, Dr. Cowan was the Chief Statistician for the Resolution Trust Corporation and the Federal Deposit Insurance Corporation, where he designed research necessary to measure the loss from the Savings & Loan Crisis of the late 1980's and capitalization requirements for the RTC funds from the U.S. Treasury.
Dr. Cowan also served as the Chief Statistician for the U.S. Department of Education, where he designed large-scale surveys of educational institutions to measure resource needs and availability, and for Opinion Research Corporation, where he designed predictive models of demand for automobile manufacturers, banks, and large horizontally diverse

firms like GE and AT&T. Dr. Cowan worked for the U.S. Bureau of the Census, where he was the Chief of the Survey Design Branch and developed many of the techniques in use today for the evaluation of coverage in surveys and censuses.

EDUCATION

Ph.D., Mathematical Statistics, The George Washington University, 1984
M.A., Economics, The University of Michigan, 1973
B.A., English and B.A., Economics, The University of Michigan, 1972

PROFESSIONAL EXPERIENCE

Co-Founder, ANALYTIC FOCUS LLC, January, 2002 to present.
Director, ARPC, November, 1999 to December, 2001.
Director, PricewaterhouseCoopers LLP, January 1997 to November, 1999.
Chief Statistician, Federal Deposit Insurance Corporation / RTC, 1991 to 1996.
Chief Statistician, Opinion Research Corporation, 1989 to 1991.
Chief Statistician, National Center for Education Statistics, US Dept. of Education, 1986 to 1989.
Bureau of the Census: Assistant Division Chief, International Statistical Programs Center, 1984 to 1986; Staff Liaison for Statistical Litigation Support, 1983 to 1984; Chief, Survey Design Branch, Statistical Methods Division, 1978 to 1983; Acting Chief, Survey Analysis and Evaluation Branch, Demographic Surveys Division, 1976 to 1978; Office of the Chief, Statistical Research Division, 1975 to 1976
Survey Research Center, Oregon State University: Manager, 1974 to 1975
Institute for Social Research, U. of Michigan: Assistant Study Director, 1972 to 1974.

PROFESSIONAL ASSOCIATIONS

Professor, Statistics, University of Alabama – Birmingham, 2002-2020. Retired.
Adjunct Professor, Harvard University, 2015 – 2016.
Associate Professor, Statistics, George Washington University, 1993 - 1998.
Visiting Research Professor, Survey Research Laboratory, U. of Illinois, 1983 - 1989.
Consultant, Dept. of Community Psychiatry, Johns Hopkins U., July 1985 - Dec 1987.

PROFESSIONAL SOCIETIES – MEMBERSHIPS

American Statistical Association (ASA)
American Association for Public Opinion Research (AAPOR)
International Association of Assessment Officers
Association for the Advancement of Wound Care

PROFESSIONAL SOCIETIES - POSITIONS

President, Research Industry Coalition, 1999-2000

Council Member, Research Industry Coalition, Representative from ASA, 1995-2000
President, Washington/Baltimore Chapter of AAPOR, 1998
Program Chair, American Association for Public Opinion Research, 1991-1992
Program Chair, Section on Survey Research Methods, ASA, 1989-90
Secretary-Treasurer, AAPOR, 1985-1986
Associate Secretary-Treasurer, AAPOR, 1984-1985
Editorial Board, Public Opinion Quarterly, 1980-1984
Editorial Board, Marketing Research, 1989-2000
Chair, Conference Committee, AAPOR, 1982-1989
Chair, Committee on Privacy and Confidentiality, ASA, 1980-1981

PUBLICATIONS

Strumpel, Burkhard; Cowan, Charles; Juster, F. Thomas; and Schmiedeskamp, Jay; editors, Surveys of Consumers 1972-73, Contributions to Behavioral Economics, Ann Arbor: The Institute for Social Research, 1975.

Duncan, Greg, and Cowan, Charles D., "Labor Market Discrimination and Nonpecuniary Work Rewards" in Surveys of Consumers 1972-73, Contributions to Behavioral Economics, Ann Arbor: The Institute for Social Research, 1975.

Curtin, Richard T. and Cowan, Charles D. "Public Attitudes Toward Fiscal Progress" in Surveys of Consumers 1972-73, Contributions to Behavioral Economics, Ann Arbor: The Institute for Social Research, 1975.

Goldfield, Edwin D.; Turner, Anthony G.; Cowan, Charles D.; Scott, John C., “Privacy and Confidentiality as Factors in Survey Response” in Public Data Use, Vol. 6, No 4, July 1978.

Cowan, Charles D., and Spoeri, Randall K., "Statistical Distance Measures and Test Site Selection: Some Considerations", Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface, 1978.

Bushery, John R., Cowan, Charles D., and Murphy, Linda R., "Experiments in Telephone-Personal Visit Surveys", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1978.

Spoeri, Randall K., and Cowan, Charles D., "On the Use of Distance Measures in Test Site Selection: A Practical Application Using Census Data", Proceedings of the American Statistical Association, Section on Business and Economic Statistics, 1978.

Cowan, Charles D.; Murphy, Linda R.; Wiener, Judy, US Bureau of the Census, “Effects of Supplemental Questions on Victimization Estimates from the National Crime Survey” in Proceedings of the American Statistical Association, Section on Survey Research Methods, 1978.

Bateman, David V.; Cowan, Charles D., US Bureau of the Census, “Plans for the 1980 Census Coverage Evaluation” in Proceedings of the American Statistical Association, Section on Survey Research Methods, 1979.

Hogan, Howard, and Cowan, Charles D., "Imputations, Response Errors, and Matching in Dual System Estimation", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1980.

Schwartz, Sidney H., Cowan, Charles D., and Sausman, Kenneth R., "Optimization in the Design of a Large-Scale State Sample", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1980.

Cowan, Charles D., "Modifications to Capture-Recapture Estimation in the Presence of Errors in the Data" presented at the meetings of the American Statistical Association, Biometrics Section, 1982 (no proceedings).

Fay, Robert; Cowan, Charles, US Bureau of the Census, “Missing Data Problems in Coverage Evaluation Studies” in Proceedings of the American Statistical Association, Section on Survey Research Methods, 1983.

Cowan, Charles D.; Fay, Robert E., “Estimates of Undercount in the 1980 Census” in Proceedings of the American Statistical Association, Section on Survey Research Methods, 1984.

Cowan, Charles D. "Interviews and Interviewing", The Social Science Encyclopedia, Routledge and Kegan Paul, Publishers, The Netherlands, 1984.

Wei, L. J. and Cowan, Charles D. "Selection Bias", Encyclopedia of Statistical Science, John Wiley and Sons, New York, N.Y., 1984.

Cowan, Charles D. and Malec, Donald J. "Capture-Recapture Models When Both Sources Have Clustered Observations", Journal of the American Statistical Association, June 1986, Vol. 81, # 394, pp. 347-353, and Proceedings of the American Statistical Association, Section on Survey Research Methods, 1984.

Cowan, Charles D., The Effects of Misclassification on Estimates from Capture-Recapture Studies. Unpublished doctoral dissertation, The George Washington University, September 1984.

Sudman, Seymour; Cowan, Charles D., "Questionnaire Design Activities in Government Statistics Offices" in Special Issue on Questionnaire Design, Journal of Official Statistics, Vol. 1, No. 2, 1985.

Cowan, Charles D. "Misclassification of Categorical Data", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1985.

Cowan, Charles D., Biemer, Paul P., Magnani, Robert J., and Turner, Anthony G., Evaluating Censuses of Population and Housing, Statistical Training Document, ISP-TR-5, U.S. Department of Commerce, Bureau of the Census, 1985.

Cowan, Charles D., Turner, Anthony G., and Stanecki, Karen "Design of the Somali Post Enumeration Survey (1986-1987)", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1986.

Cowan, Charles D., Breakey, William R., and Fischer, Pamela J. "The Methodology of Counting the Homeless", Proceedings of the American Statistical Association, Section on Survey Research Methods, 1986.

Cowan, Charles D. and Malec, Donald J. "Sample Allocation for a Multistage, Multilevel, Multivariate Survey", Proceedings of the Fourth Annual Research Conference (ARC IV), U.S. Bureau of the Census, 1988.

Frey, Carolin M., McMillen, Marilyn M., Cowan, Charles D., Horm, John W., and Kessler, Larry G.. "Representativeness of the Surveillance, Epidemiology, and End Results Program Data: Recent Trends in Mortality Rates", Journal of the National Cancer Institute, Vol. 84, No. 11, June 3, 1992.

Cowan, Charles D., Breakey, William R., and Fischer, Pamela J. "The Methodology of Counting the Homeless, A Review" in Homelessness, Health, and Human Needs. Institute of Medicine, National Academy Press, National Academy of Sciences, Washington, D.C., 1988.

Cowan, Charles D., "Standards for Statistical Surveys in the Federal Government: Practices in the Center for Education Statistics", Proceedings of the American Statistical Association, Section on Survey Methods Research, 1988.

Sudman, Seymour, Sirken, Monroe G., and Cowan, Charles D., "Sampling Rare and Elusive Populations", Science, Vol. 240, pp. 991-996, May 20, 1988.

Cowan, Charles D., "Mall Intercepts and Clinical Trials: The Philosophy of Inference from Different Types of Research Designs" in Marketing Research: A Magazine of Management & Applications, Vol. 1, No. 1, March 1989.

Cowan, Charles D., "Mall Intercepts: Principles of Design for Research" in Proceedings of the Seventh Annual Advertising Research Foundation Research Quality Workshop, September 1989.

Cowan, Charles D., "Estimating Census and Survey Undercounts Through Multiple Service Contacts" in Housing Policy Debate: Counting the Homeless: The Methodologies, Policies, and Social Significance Behind the Numbers, Volume 2, Issue 3, pp. 869-882, 1991.

Cowan, Charles D., "Ratio vs. Regression Estimators in a Large Scale Survey of S&L's" in Proceedings of the Section on Survey Research Methods, American Statistical Association, 1992.

Cowan, Charles D., "A Longitudinal Survey and Reality Check for the Value of Financial Assets" in Proceedings of Statistics Canada Symposium 92: Design and Analysis of Longitudinal Surveys, November 1992.

Cowan, Charles D., and Wittes, Janet, "Intercept Studies, Clinical Trials, and Cluster Experiments: To Whom Can We Extrapolate?" in Controlled Clinical Trials, Vol.15, pp.24-29, 1994.

Cowan, Charles D.; Klena, Mathew J., Resolution Trust Corp, “Allocation of Proceeds from Bulk Auctions to Individual Assets” in Proceedings of the American Statistical Association, Section on Business and Economic Statistics, 1995.

Cowan, Charles D. "Statistical Sampling as a Management Tool in Banking" in FDIC Banking Review, 1997, Vol. 10, No. 1.

Cowan, Charles D., “Coverage, Sample Design, and Weighting in Three Federal Surveys” in Journal of Drug Issues, October 2001.

Cowan, Charles D., "Use of Mass Appraisals in Toxic Tort Litigation Involving Loss of Value" in Proceedings of the International Association of Assessment Officers, October 2002.

Cowan, Adrian M. and Cowan, Charles D., "Default Correlation: An Empirical Investigation of a Subprime Lender", The Journal of Banking and Finance, March 2004.

Cowan, Charles D. and Cowan, Adrian M., "A Survey Based Assessment of Financial Institution Use of Credit Scoring for Small Business Lending", SBA Report 283, Nov. 2006

Keith, Scott W., Wang, Chenxi, Fontaine, Kevin R., Cowan, Charles D. and Allison, David B. "Body Mass Index and Headache Among Women: Results From 11 Epidemiologic Datasets", Obesity, Volume 16, Issue 2 (February 2008) 16: 377-383; doi:10.1038/oby.2007.32

Cowan, Adrian M. and Cowan, Charles D., "The Dynamics of Credit Quality and Implications for the Pricing of Small Business Loans", The International Journal of Banking and Finance, 2007/08 (March) Vol. 5. Number 2:2008: 31-60

Brock, David W., Thomas, Olivia, Cowan, Charles D., Hunter, Gary R., Gaesser, Glenn A., and Allison, David B., Association between Physical Inactivity and Prevalence of Obesity in the United States, Journal of Physical Activity and Health, January, 2009

Cowan, Charles D., "Use of Statistical Analysis to Measure Damages", in Comprehensive Guide to Economic Damages, Fourth Edition, Business Valuation Resources, LLC, Portland, OR, April, 2016.

Cowan, Charles D., "Qui Custodiet Custodiens?", International In-House Counsel Journal, Cambridge, England, October 20.

XI. Exhibit 2: Testimony in Last Four Years

Financial:

Homeward Residential as Master Servicer for Option One Mortgage Loan Trust 2006-2 v. Sand Canyon Corporation, f/k/a Option One Mortgage Corporation. Worked for plaintiff. Deposed in January 2015; deposed again in September 2015; hearing, June 2017.

Western Southern Life Insurance Company v. Bank of New York Mellon. Worked for plaintiff. Deposition, January 2016. Testified at trial, January 2017.

MBIA v Credit Suisse. Worked for plaintiff. Deposed in January 2016. Trial in July 2019.

FHFA v. RBS. Worked for plaintiff. Deposed in October 2016.

Mass Mutual v. Credit Suisse First Boston Mortgage Securities Corp. et al. Worked for plaintiff. Deposed in November 2016. Trial in August 2017.

FDIC as Receiver for Colonial Bank v. Credit Suisse First Boston Mortgage Securities Corp.; Credit Suisse Management LLC; Credit Suisse Securities (USA) LLC; First Horizon Asset Securities, Inc.; First Horizon Home Loan Corporation; FTN Financial Securities Corp.; and HSBC Securities (USA) Inc.; FDIC as Receiver for Receiver for Colonial Bank v RBS Securities Inc. Worked for plaintiff. Deposed in March 2017.

Mass Mutual v. Goldman Sachs et al.. Worked for plaintiff. Deposed in May 2017.

CUNA v. Morgan Stanley. Worked for plaintiff. Deposed in May 2017.

CUNA v. Credit Suisse. Worked for plaintiff. Deposed in July 2017. Deposed on rebuttal report, September 2017.

FHLB – Chicago v. Morgan Stanley. Worked for plaintiff. Deposed in October 2017.

FDIC as Receiver for Guaranty Bank v RBS Securities Inc. Worked for plaintiff. Deposed in November 2017.

People of the State of California v Morgan Stanley et al., Worked for plaintiff. Deposed in January 2018.

FDIC as Receiver for Receiver for Guaranty Bank v Goldman, Sachs & Co. and Deutsche Bank Securities. Worked for plaintiff. Deposed in January 2018.

Christopher S. Porrino, Attorney General of New Jersey on behalf of Amy G. Kopleton, Deputy Chief of the New Jersey Bureau of Securities, v. Credit Suisse Securities (USA) LLC, Credit Suisse First Boston Mortgage Securities Corp., and DLJ Mortgage Capital, Inc., Worked for plaintiff. Deposed in November 2018.

Federal Home Loan Bank of Boston v. Nomura; Federal Home Loan Bank of Boston v. Credit Suisse Securities (USA) LLC, Credit Suisse First Boston Mortgage Securities Corp.,. Worked for plaintiff. Deposed in January 2019, day 1, February 2019, day 2.

Financial Guaranty Insurance Company v. Morgan Stanley ABS Capital I Inc. and Morgan Stanley Mortgage Capital Holdings LLC, as successor to Morgan Stanley Mortgage Capital Inc., Worked for plaintiff. Deposed in May 2019.

AMBAC Assurance Corporation et al v. First Franklin et al. Worked for plaintiff. Deposed in December 2019.

Financial - non RMBS

In Re Riddell Concussion Reduction Litigation. Worked for plaintiff. Deposed Jan. 2017.

IPOX Schuster, LLC v. Nikko Asset Worked for defendants. Deposed in July 2017.

Charles Baird and Lauren Slayton, as individuals, and on behalf of all others similarly situated, and on behalf of the BlackRock Retirement Savings Plan v. BlackRock Institutional Trust Company, N.A.; BlackRock, Inc.; The BlackRock, Inc. Retirement Committee; The Investment Committee of the Retirement Committee; The Administrative Committee of the Retirement Committee; The Management Development & Compensation Committee, Catherine Bolz, Chip Castille, Paige Dickow, Daniel A. Dunay, Jeffrey A. Smith; Anne Ackerley, Amy Engel, Nancy Everett, Joseph Feliciani Jr., Ann Marie Petach, Michael Fredericks, Corin Frost, Daniel Gamba, Kevin Holt, Chris Jones, Philippe Matsumoto, John Perlowski, Andy Phillips, Kurt Schansinger, Tom Skrobe; Kathleen Nedl, Marc Comerchero, Joel Davies, John Davis, Milan Lint, Laraine McKinnon, and Mercer Investment Consulting. Worked for plaintiffs. Deposed in May 2019.

Disparate Impact \ Discrimination:

River Cross Land Company, LLC v. Seminole County, FL. Worked for plaintiff. Deposition, October 2019.

Construction Defects:

Abad v. Western Pacific Housing. Worked for defense. Deposition, October 2016.

Stanton v. Richmond American Homes of Nevada Inc. Worked for defense. Deposition, October 2016.

Park et al. v. Meritage Homes. Worked for defense. Deposition, December 2017.

Vizcayne North Condominium Association, Inc. v W.G. Yates & Sons Construction Company, et al. (also South Condominium Association and Master). Worked for defense. Deposition, February 2017.

Donald Melosh, et al. v. Western Pacific Housing, Inc., JAMS Case No.: 1100091610, Construction Defects. Worked for defense. Deposition, March, 2020.

Other Cases:

Estate of Evan Press v Northport Health Services of Arkansas, LLC and NHS Management, LLC. Wrongful Injury Suit in Alabama. Worked for defendants. Deposition, July 2017.

Pudlowski v. St. Louis Rams. Worked for plaintiff. Deposed in September 2017. Deposed again in October 2018.

Elena Tyurina v. Urbana Tahoe TC LLC, Urbana Tahoe Beverage Company, LLC dba Beach Retreat and Lodge Tahoe, and Action Motorsports of Tahoe, Inc. Worked for Defendant. Deposition, April 2018.

In Re: Dicamba Herbicides Litigation. Product Defect case. Worked for plaintiffs. Deposition, March 2019.

Otter Products et al, v. Phone Rehab et al. Deceptive Sales. Worked for plaintiff. Deposed in November 2019.

Thomas Allegra et al v. Luxottica Retail North America. Worked for plaintiff. Deposed in December 2019.

Westgate Resorts v. Reid Hein & Associates, dba Timeshare Exit Team. Tortious Interference case. Worked for plaintiffs. Deposition, March 2020.

Charles Copley et al Bactolac Pharmaceutical, Inc. et al. Worked for Plaintiffs, Deposed August 2020.

Mildred Clemmons et al v. Samsung Electronics of America, Inc. Worked for plaintiffs. Deposed in October 2020.

XII. Exhibit 3: Documents Relied Upon

In addition to my analyses, I relied on the following documents for my report:

1. Third Amended Consolidated Complaint, ECF No. 166
2. Kendall's Advanced Theory of Statistics, Wiley, London, 1998, 6th Edition, Volume 1, by Alan Stuart & Keith Ord (Monte Carlo method)
3. Facebook's 2015-2020 10-Q and 10-K filings with the U.S. Securities and Exchange Commission
4. Facebook Q4 2019 Results, available at https://s21.q4cdn.com/399680738/files/doc_financials/2019/q4/Q4-2019-Earnings-Presentation-_final.pdf
5. Aaron Smith and Monica Anderson, "Social Media Use in 2018: Demographics and Statistics." Pew Research Center, March 1, 2018. Available at: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/
6. FB-SINGER-00258878
7. FB-SINGER-00184892
8. FB-SINGER-00170105
9. FB-SINGER-00022619
10. FB-SINGER-00258878
11. FB-SINGER-00169905
12. FB-SINGER-00315234
13. FB-SINGER-00426324
14. FB-SINGER-00304046
15. FB-SINGER-00304054
16. FB-SINGER-00304060
17. FB-SINGER-00426302
18. FB-SINGER-00426314

19. FB-SINGER-00426321
20. FB-SINGER-00426295
21. FB-SINGER-00314173
22. FB-SINGER-035
23. FB-SINGER-026
24. FB-SINGER-041
25. FB-SINGER-00084221
26. FB-SINGER-00313499
27. FB-SINGER-00080023
28. FB-SINGER-0025712 (Plaintiff's Deposition Exhibit 73)
29. FB-SINGER-00164997 (Plaintiff's Deposition Exhibit 34)
30. FB-SINGER-00169905 (Plaintiff's Deposition Exhibit 72)
31. FB-SINGER-00331431
32. FB-SINGER-00255393
33. FB-SINGER-00264321
34. FB-SINGER-00183123
35. Facebook, Inc.'s Third Amended And Supplemental Responses And Objections To Plaintiffs' Second Set Of Interrogatories (Plaintiffs' Exhibit 134)
36. Christopher Ingraham, "Map: The coronavirus risk for a big Thanksgiving dinner in your community." *Washington Post,* November. 18, 2020. Available at: https://www.washingtonpost.com/business/2020/11/18/map-covid-risk-thanksgiving/
37. Brittany Shammas and Lena H. Sun, "How the Sturgis Motorcycle Rally may have spread coronavirus across the Upper Midwest." *Washington Post*, October 17, 2020. Available at: https://www.washingtonpost.com/health/2020/10/17/sturgis-rally-spread/
38. COVID-19 Event Risk Assessment Planning Tool, available at: https://covid19risk.biosci.gatech.edu

39. Jack Hayya, Donald Armstrong, Nicolas Gressis, (1975). "A Note on the Ratio of Two Normally Distributed Variables." *Management Science* 21(11):1338-1341, available at: https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.21.11.1338.

40. T. Pham-Gia, N. Turkkan & E. Marchand (2006). "Density of the Ratio of Two Normal Random Variables and Applications." *Communications in Statistics - Theory and Methods*, 35:9, 1569-1591, DOI: 10.1080/03610920600683689, available at: https://www.researchgate.net/publication/233479623_Density_of_the_Ratio_of_Two_Normal_Random_Variables_and_Applications

41. Hinkley, D. (1969). "On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 56(3), 635-639. doi:10.2307/2334671, available at: http://static.stevereads.com/papers_to_read/on_the_ratio_of_two_correlated_normal_random_variables.pdf

42. Hinkley, D. (1970). "Correction: On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 57(3), 683-683. doi:10.2307/2334796, available at: http://static.stevereads.com/papers_to_read/correction_on_the_ratio_of_two_correlated_normal_random_variables.pdf

43. George Marsaglia (April 1964). "Ratios of Normal Variables and Ratios of Sums of Uniform Variables." *Defense Technical Information Center,* available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/600972.pdf

44. Marsaglia, G. (2006)." Ratios of Normal Variables. *Journal of Statistical Software*. 16(4), 1 - 10. doi:10.18637. Available at: https://www.jstatsoft.org/article/view/v016i04

45. Transcript of Yaron Fidler Deposition

46. Transcript of David Amsallem Deposition

47. Transcript of FRCP 30(b)(6) Gerardo Zaragoza Deposition

48. Transcript of FRCP 30(b)(6) Rahul Bhandari Deposition

49. Expert Report of Dr. Atif Hashmi

XIII. APPENDICES

189. Appendices 1, 5, 7, 11, 12, 13, and 14 follow herein.

190. Appendices 2 (Tabs A & B), 3, 4, 6, 8, 9, 10, and 15 are native Excel files, served separately along with other worksheets and analyses relied on for this report.

A. APPENDIX 1: Sensitivity Analysis – US Default Potential Reach Calculation

191. In this Appendix, I perform a sensitivity analysis to show the range of overlap from five sources of inflation (i.e. the maximum and minimum) in calculating the initial default Potential Reach inflation for the US. The purpose of this analysis is to show that the calculation I perform in Section VII.A assumes a reasonable overlap among the five sources of inflation, given the possible range of overlap.

192. As shown in Table 6 of the report, there is a 45% overlap between Facebook and Instagram accounts, and as shown in Section VII.A.4, [redacted] of the cross-platform overlap is unaccounted for by Facebook. Thus, only [redacted] of cross-platform accounts are deduplicated. Based on that and the information in previous sections, the following values are necessary for calculating the effect of cross-platform inflation.

Table 12: Values Needed to Compute Inflation from All 5 Sources for US Only

Inactive	8.08%		91.92%	Active
Ineligible	[redacted]		[redacted]	Eligible
Fake	5.00%		95.00%	Genuine
% Two Accounts (SUMA)	[redacted]		[redacted]	% One Account (SUSA)
Unaccounted Overlap	[redacted]		[redacted]	Accounted for Overlap
Overlap of FB and IG	45.00%		55.00%	No Overlap of IG and FB
SUMA	[redacted]		[redacted]	Product of Five sources of inflation

193. Tables 13 and 14 below expand on Table 9 in Section VII.A.9. But while 45% of accounts are affected by the overlap between Facebook and Instagram accounts, 21 % of

those are unaccounted for, i.e. there are cross-platform duplicates that Facebook did not deduplicate.

194. In Table 13, I first determine what the possible overlap can be between all five sources of inflation, assuming they are independent, i.e., the probability that any given account is Genuine, Active, Eligible, SUSA, and is not a cross-platform duplicate (I will use "corresponds to a person" as shorthand). I determine that the overlap is 59.15%. This is calculated as follows: Active (.9192) x Eligible [redacted] x Genuine (.95) x SUSA [redacted] x Accounted for in Cross-Platform Overlap [redacted] .5915 or 59.15%.

Table 13: Overlap of FB and IG, 5 Variables, No Duplicates in Cross-Platform Overlap

% SUSA, One Account

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	**59.15%**	5.33%	64.5%
No	5.20%	0.47%	5.7%
Total	64.3%	5.8%	70.1%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	3.11%	0.28%	3.4%
No	0.27%	0.02%	0.3%
Total	3.4%	0.31%	3.7%

% SUMA, Two Accounts

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	4.14%	0.37%	4.5%
No	0.36%	0.03%	0.4%
Total	4.5%	0.4%	4.9%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	0.22%	0.02%	0.24%
No	0.08%	0.00%	0.02%
Total	0.24%	0.02%	0.26%

Minimum Overlap

Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**58.68%**	5.80%	64.5%
No	5.67%	0%	5.7%
Total	64.3%	5.8%	70.1%

Maximum Overlap

Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**64.34%**	0.13%	64.5%
No	0%	5.67%	5.7%
Total	64.3%	5.8%	70.1%

195. If I assume that all five sources of inflation are not independent, but correlated, based on the numbers Facebook provides, there will be a range of possible overlap between them. Given the minimum possible overlap, the likelihood that a given account corresponds to a person is 58.68% (59.15% – 0.47%). Under the maximum possible overlap, that likelihood is 64.34% (59.146% + 5.197%). Accordingly, 59.15% is a reasonable expected percentage that lies within the range (58.68% to 64.34%) set by the minimum and maximum possible overlap.

196. Next, in Table 14, I account for the [redacted] of the accounts affected by the cross-platform overlap that are not deduplicated.

Table 14: Overlap of FB and IG, 5 Variables, Cross-Platform Duplicates in Overlap

% SUSA, One Account

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	**15.72%**	1.42%	17.1%
No	1.38%	0.12%	1.5%
Total	17.1%	1.5%	18.6%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	0.83%	0.07%	0.9%
No	0.07%	0.01%	0.1%
Total	0.9%	0.08%	1.0%

% SUMA, Two Accounts

Genuine	*Eligible*		
Active	Yes	No	Total
Yes	1.10%	0.10%	1.2%
No	0.10%	0.01%	0.1%
Total	1.2%	0.1%	1.3%

Fake	*Eligible*		
Active	Yes	No	Total
Yes	0.06%	0.01%	0.06%
No	0.18%	0.00%	0.01%
Total	0.06%	0.01%	0.07%

Minimum Overlap			
Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**15.60%**	1.54%	17.1%
No	1.51%	0%	1.5%
Total	17.1%	1.5%	18.6%

Maximum Overlap			
Genuine & SUSA	*Eligible*		
Active	Yes	No	Total
Yes	**17.10%**	0.03%	17.1%
No	0%	1.51%	1.5%
Total	17.1%	1.5%	18.6%

197. For that [redacted] f the 45% of accounts affected by overlap between Facebook and Instagram, the likelihood that any given account corresponds to a person is 15.72%. This was calculated as follows: Active (.9192) x Eligible ([redacted]) x Genuine (.95) x SUSA [redacted] Unaccounted Duplicates in Cross-platform Overlap [redacted] = .1572

198. If I assume the minimum possible overlap, the likelihood that any given account corresponds to a person is 15.60% (15.72% - .12%). If I assume the maximum possible overlap, that likelihood becomes 17.10% (15.72% + 1.38%).

199. I assume 15.72% is a reasonable expected likelihood because it is within the narrow range of permissible probabilities (15.60% to 17.10%). Assuming a 15.72% likelihood, the inflation rate for the [redacted] of the 45% of accounts that were not de-duplicated across platforms is 1.592 or 59.2%. This is calculated as: (SUMA [redacted] Unaccounted Duplicates in Cross-platform Overlap [redacted] / (Active (.9192) x Eligible [redacted] x Genuine (.95)) = 1.592

200. Calculations for the ranges of overlap between five sources of inflation, using US-only rates are also in Appendix 4.

B. APPENDIX 5: SUMA for Narrower Targeting and Multiple SUMA Accounts

201. In this Appendix, I discuss two considerations when determining the SUMA rates for this report based on Facebook's documents.

202. The first is FB-SINGER-00314173, which (at '197) appears to show a much higher SUMA rate due to users that have more than one duplicate account. In FB-SINGER-00314173 Facebook reports a very different distribution of accounts, with some people having as many as six accounts in research Facebook conducted. The resulting value of SUMA using the numbers reported [redacted] which is a SUMA value far exceeding that in any of the other Facebook documents.

Table 15: Distribution of SUMA Accounts Based on FB-SINGER-00314173 at '197

[redacted]

203. One potential explanation for this is that, instead of computing the SUMA rate using the number of accounts, Facebook collapses the multiple accounts into only 2 accounts, regardless of the number of accounts a person may actually have. If Facebook collapses the number of accounts to 2, then the SUMA value that would be calculated

would drop to [redacted] This number is still much larger than most of the values reported by Facebook, but this procedure would explain, in part, the disparity between the SUMA numbers reported elsewhere and the numbers computed from this document.

204. However, to be conservative and use the most robust and reliable data available, I did not use FB-SINGER-00314173 for SUMA rates in my report, but instead used FB-SINGER-00426295.

205. The second consideration is the fact that Facebook did build a model for removing SUMA ("SUMA model") from Potential Reach that would account for more granular targeting criteria. However, Facebook only validated the SUMA model at the country and age level.[69]

206. At his deposition, Facebook employee David Amsallem testified that:

> In this case the SUMA model I worked with is a model built by the growth team to estimate at country level what -- what SUMA rates are. So what percentage accounts may belong to the same person.
>
> This, the SUMA model, I explored what would be the impact on the reach estimates if we were to apply it? And that being said, this model wasn't validated for that purpose **because it was validated for the purpose of understanding at a country level what SUMA rates are**. And the precision of the model was an introduct [sic] use case.
>
> **Once you go into targeting, you have millions of different options, and so you need actually to test for these different use cases whether the model is still valid or not**, whether it may predict that you and I are the same person even though it's not the case. [70]

207. Despite the lack of validation for all targeting specifications, the Product Manager for Potential Reach advocated using the SUMA model for Potential Reach.[71] And, by

---

[69] FB-SINGER-00025712 (Plaintiffs' Deposition Exhibit 73)

[70] David Amsallem Deposition, at 69:17-70:12 (emphasis added).

[71] FB-SINGER-00164997, at '98 ("…IMHO we should start use the data s is [sic] since it's the best we have and improve it over time.")

December 2018, the growth identity team also confirmed that it was comfortable with ads using the SUMA model.[72]

208. In the absence of Facebook having ever validated the SUMA model for targeting specifications other than country and age, as I discussed in Section VI.G, I can use statistical analyses to determine the range of SUMA inflation for each country/age combination even if the advertiser also included additional types of targeting.

209. I have also reviewed a document where Facebook appears to have used the SUMA model to calculate the Potential Reach for additional types of targeting criteria.[73] However, because the SUMA model was not validated for these types of targeting criteria, these documents do not reflect the proper approach for evaluating the impact of SUMA accounts on these additional types of targeting criteria. Instead, the appropriate approach is to use the most robust and reliable distribution of SUMA available, which, as discussed above in Sections VI.A and VI.B, is based on FB-SINGER-00426295.

210. Furthermore, ultimately, as I show in my sensitivity analysis in Appendix 11, even if the SUMA rate went all the way down to 0% (or 1.0), there would be inflation well above 10%, due to all the other existing sources of inflation.

C. APPENDIX 7: Ineligible Account Rates Regression

211. In the regression, I recognize that the Ineligible rate averaged across the 30 countries is only [redacted] whereas the global Ineligible rate provided by Facebook in discovery is [redacted] did not correct these values so that they would average to the published rate, to the published rate, so my result is more conservative in that the relationship between the Eligible rate and the SUMA rate found in the 30 countries is not the same on average in other countries not included in the regression. The relationship shows a declining

---

[72] FB-SINGER-00169905; David Amsallem Deposition, at 183:17-190:8.

[73] FB-SINGER-00331431.

Eligible rate as SUMA increases, to increase the variability of my results and make them much more conservative for countries estimated to have higher rates of Eligible accounts.

212. I then ran a regression between the Eligible rate and the SUMA rate plus indicator variables for each of the age groups. After running the regression, I examined the originals and found a single observation, for Spain, age 13-17, that was a very extreme outlier. I removed this observation and ran the regression a second time and found other outliers. I define outliers as observations whose absolute value standardized residual was greater than 3.291, which would make it outside a confidence interval of 99.9%. That is, the residual would have only a 1 in 1000 chance of being from the model being fit in the regression.

213. This led me to eliminate more observations from the refit regression, but at this point I also observed in examining the residuals that the residuals were especially large for Russia, Korea, and Japan. I added indicator variables to the regression for these three countries to account for the variability due to these countries. This increased the amount of variability explained in the eligibility rate from about 20% to over 60% (as measured by R-squared). Further tests of the residuals removed a total of five observations after the addition of the country indicator variables. As a result, the number of observations dropped from 330 to 325, a modest loss of five observations that had the potential to skew the regression.

214. I forecast the Eligible rate for every other country except the 30 countries provided in the discovery for Facebook. For the 30 countries I used exactly the values provided by Facebook even though some of them seem extreme. For the remaining 210+ countries and age groups I use the values estimated from the regression. The final regression is presented in the backup materials as well as the buildup to these results.

D. APPENDIX 11: Sensitivity Analysis – Excluding Sources of Inflation

215. In the analyses performed in this report, I rely on four factors that are sources of inflation: SUMA, Ineligible, Inactive, and Fake accounts. In this sensitivity analysis, rather than recalculating the Potential Reach inflation for a range of lower values for SUMA, Ineligible, Inactive, and Fake account rates, I will make an even more conservative assumption and calculate the Potential Reach inflation assuming that any one of those inflation sources did not exist. I considered each of these scenarios and compared each to the inflation computation using the four sources. The values with one factor removed are computed by taking the full inflation and dividing or multiplying by the appropriate factor. [74] The results are in Table 16 below.

Table 16: Potential Reach Inflation When the Rate of Any One of Four Sources of Inflation is 1.0 or 0%

	SUMA = 1	Eligible = 1	Active = 1	Genuine = 1	Full Inflation
Inflation =	1.3370	1.2675	1.3603	1.4059	1.4799

216. I performed a similar analysis to observe the effect on inflation in the US default Potential Reach, if any of the five sources of inflation did not exist. The results are in Table 17 below.

Table 17: US Default Potential Reach Inflation When the Rate of Any One of Five Sources of Inflation is 1.0 or 0%

	SUMA = 1	Eligible = 1	Active = 1	Genuine = 1	Total Inflation	
Inflation =	1.3067	1.2772	1.2798	1.3226	1.3923	**Without IG**
Inflation =	1.3662	1.3253	1.3280	1.3724	1.4447	**With IG**

[74] *See also* Appendix 2B.

217. Thus, for instance, if I excluded Ineligible accounts by changing the Ineligible rate to 0%, the US default Potential Reach would still be inflated by 32.53%. [75]

218. The conclusion from this sensitivity analysis is that, even if I go to the extreme of eliminating one of the sources of inflation completely from the computation, my opinion that Potential Reach inflation of at least 10% is a statistical certainty remains unchanged.

E. APPENDIX 12: Sensitivity Analysis – Assuming Normal Distribution of Potential Reach Inflation Rates

219. Throughout my analysis, to compute probabilities I have assumed that the Potential Reach inflation rate is normally distributed (the bell curve). In the Monte Carlo simulations in Section VII.B.1, the computation of inflation is a sum of values divided by another sum of values. The numerator is the sum of accounts that would be shown an advertisement including second accounts (SUMA). The denominator is the sum of accounts that would be genuine, eligible, and active. From the Central Limit Theorem, I know that sums tend to normality as the size of the sample (Potential Reach in this case) increases. From the academic literature, I know that the ratio of normally distributed variables is also normal, except when both the numerator and denominator have a mean of zero (in which case the ratio is distributed as a Cauchy distribution)[76]. Therefore, inflation can be considered a normally distributed variable.

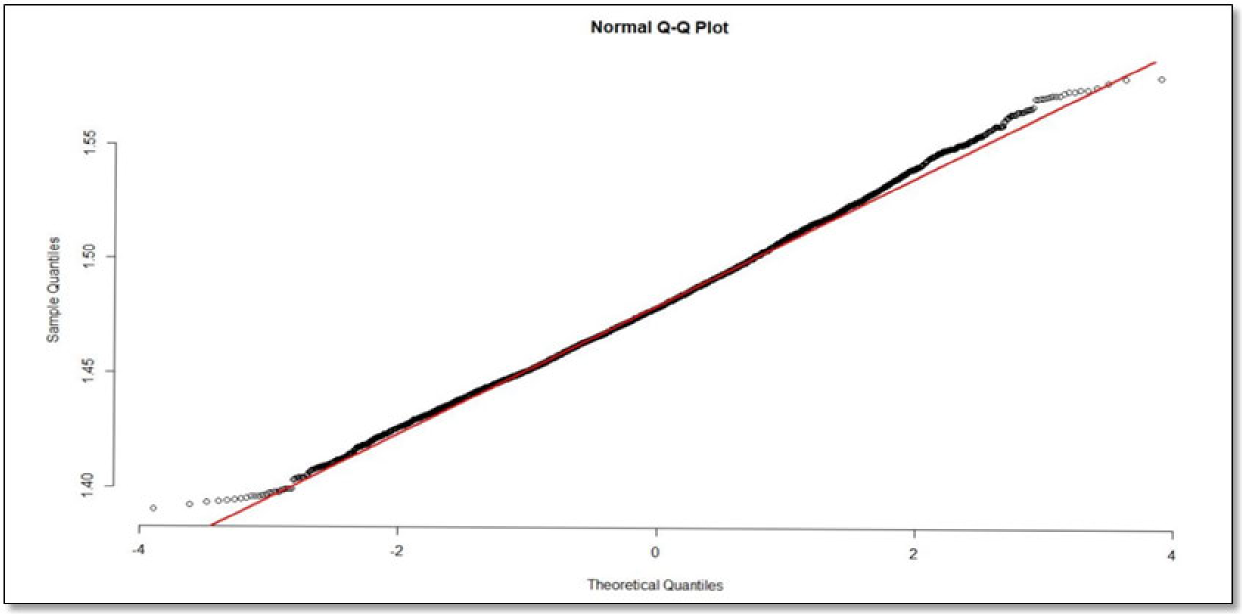
---

[75] My understanding is that the SUMA and Fake account rates have gone up over time, according to Facebook's own SEC filings. Therefore, these calculations are conservative.

[76] Jack Hayya, Donald Armstrong, Nicolas Gressis, (1975). "A Note on the Ratio of Two Normally Distributed Variables." *Management Science* 21(11):1338-1341, available at: https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.21.11.1338. T. Pham-Gia, N. Turkkan & E. Marchand (2006). "Density of the Ratio of Two Normal Random Variables and Applications." *Communications in Statistics - Theory and Methods*, 35:9, 1569-1591, DOI: 10.1080/03610920600683689, available at: https://www.researchgate.net/publication/233479623_Density_of_the_Ratio_of_Two_Normal_Ran

220. In an abundance of caution, I also ran standard check on normality on the 10,000 outcomes from the Monte Carlo simulation. Results are presented in Chart 5 below.

Chart 5: Outcomes of Tests to Determine if Inflation Results from the Monte Carlo Simulation Follow the Normal Distribution



221. Notably, the software from SAS and Stata do not run some tests for normality because the sample size is very large, exceeding the limits allowed by the software. However, it is always possible to plot a Q-Q plot, which is a chart of the quantiles of the theoretical normal distribution on the horizontal axis by the quantiles for the frequency

---

dom Variables and Applications; Hinkley, D. (1969). "On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 56(3), 635-639. doi:10.2307/2334671, available at: http://static.stevereads.com/papers_to_read/on_the_ratio_of_two_correlated_normal_random_variables.pdf; Hinkley, D. (1970). "Correction: On the Ratio of Two Correlated Normal Random Variables." *Biometrika.* 57(3), 683-683. doi:10.2307/2334796, available at: http://static.stevereads.com/papers_to_read/correction_on_the_ratio_of_two_correlated_normal_random_variables.pdf; George Marsaglia (April 1964). "Ratios of Normal Variables and Ratios of Sums of Uniform Variables." *Defense Technical Information Center,* available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/600972.pdf; Marsaglia, G. (2006)." Ratios of Normal Variables. Journal of Statistical Software," 16(4), 1 - 10. doi:http://dx.doi.org/10.18637/jss.v016.i04. Available at: https://www.jstatsoft.org/article/view/v016i04

distribution of the observed data on the vertical axis. This is charted in Chart 5, which shows that the data tracks almost exactly with the normal distribution.

F. APPENDIX 13: Inflation With and Without Instagram Accounts

222. In Section VII.A, I analyze the effect of a fifth source of inflation: duplicate cross-platform accounts. The purpose of this sensitivity analysis is to demonstrate that if I make the assumption that Instagram accounts have duplicates in the same manner as do Facebook accounts,[77] the Potential Reach inflation is always greater in computations that include Instagram accounts than in computations that consider Facebook accounts by themselves.

223. Here I consider if there is ever a time that Potential Reach inflation would be lower when I consider Instagram ("IG") accounts in the computation. The answer is no.

224. The two equations presented earlier in the report for inflation are:

$$\text{Inflation with IG} = \frac{[(\%\text{FB}) * \text{SUMA} + (\%\text{FB\&IG}) * \%DedupOverlap + (\%\text{IG}{\sim}\text{FB}) * 1]}{(1-f) * (1-g) * (1-v)}$$

$$\text{Inflation without IG} = \frac{SUMA}{(1-f) * (1-g) * (1-v)}$$

I set up these two equations to test if they are equal, or one is larger than another, as in

Inflation with IG (<, =, >) Inflation without IG

I note that the denominators are the same and so cancel, and I get

$$[(\%\text{FB}) * \text{SUMA} + (\%\text{FB\&IG}) * \%DedupOverlap + (\%\text{IG}{\sim}\text{FB}) * 1] \ ? \ SUMA$$

Where the question mark denotes that I don't know whether it is a (<, =, or >) relationship.

[77] Facebook documents indicate that multiple/duplicate accounts are actually even more common on Instagram than on Facebook. *See, e.g.,* FB-SINGER-00255393 at '96.

225. I note that %FB +%IG~FB = 1.0 by the way I set up the account distributions. Therefore, if (%IG~FB) was multiplied by SUMA instead of 1.0 (which assumed SUMA was zero for Instagram accounts), I would have

$$(\%\mathrm{FB\&IG}) * \%DedupOverlap > 0$$

226. As both percentages are positive, and the product of the percentages is positive, the left side is always greater than zero. This means that if people with Instagram accounts have multiple accounts at about the same rate as people with Facebook accounts, then inflation with Instagram accounts is always greater than inflation without Instagram accounts. In the extreme, if people with Instagram accounts have no duplicate Instagram-only accounts then

$$\frac{[(\%\mathrm{FB}) * \mathrm{SUMA} + (\%\mathrm{FB\&IG}) * \%DedupOverlap + (\%\mathrm{IG{\sim}FB}) * 1]}{SUMA} \ ?\ 1.0$$

Or

$$(\%FB\&IG) * \frac{\%DedupOverlap}{SUMA} - \%IG{\sim}FB * \left(1 - \frac{1}{SUMA}\right) \ ?\ 0$$

227. The left-hand term for the U.S. is about 4%, while the right-hand term is about 13% divided by 1.1, or approximately 11.8%. The difference is significantly larger than zero for the U.S. The balancing point in this equation (where the two sides are equal) would be found only in countries where people hold a large percentage of Instagram accounts and no Facebook accounts. This increases the left side of the equation and reduces the right side of the equation until the left side is approximately zero. However, this would occur at approximately the point where there are 50% or more accounts that are Instagram only and not Facebook. However, this would be the outcome only where the Instagram accounts have no duplicates (SUMA =1), which seems highly unlikely.

G. APPENDIX 14: Potential Reach Rounding

228. It is my understanding that Facebook rounds its Potential Reach numbers, and does so to only the first two digits of the numbers.[78] This means that, at times, Facebook is rounding the Potential Reach number up, further inflating the Potential Reach value. At other times, Facebook rounds the Potential Reach number down, deflating the value. However, as I explain below this deflation can never be greater than just under 5%. Thus, any inflation at or above 5% could not be ameliorated by rounding.

229. Any number can be expressed in scientific notation, such as 13,289 = 1.3289 x 10^4. Using scientific notation, I can express actual Potential Reach as a number in the form (x.y+z) x 10^k. A rounded Potential Reach would be the same number but with the last part rounded off, which would be written as (x.y) x 10^k. The value k as an exponent of ten simply indicates whether the Potential Reach is in the hundreds (k=2), thousands (k=3), and so on. The value z is defined to be any number less than 0.1, although I explain below that z must be in the range (-0.05 to$+.04\bar{9}$) since otherwise rounding would not occur. In other words, rounding would only occur for values in the range in the range of -0.05 to$+.04\bar{9}$.

230. To determine the inflation or deflation of Potential Reach due to rounding, I divide the actual Potential Reach by the rounded Potential Reach, which is expressed as:

$$\frac{Actual\ PR}{Rounded\ PR} = \frac{(x.y+z)*10^k}{(x.y)*10^k} = 1 + \frac{z}{x.y}$$

231. Thus, z divided by x.y is the inflation or deflation in the Potential Reach, regardless of the size of the number in terms of hundreds, thousands, millions, billions, or any other

---

[78] Expert Report of Dr. Atif Hashmi at ¶ 45; *see also* Rahul Bhandari Deposition, at 67:21–68:4, FB-SINGER-00084221, at '22.

magnitude. This is because the value 10^k is in the numerator and the denominator, so the two cancel each other out.

232. The smallest that z can be is -0.05, since this value, and larger numbers up to zero, would cause the Potential Reach to be rounded up to the next value, which can be determined using the following formula: x.w x 10^k, where w is y+1.

233. The largest that z can be is $.04\overline{9}$, where the bar over the 9 indicates it repeats out to infinity. At z = .05, the Potential Reach would round up, as noted in the previous paragraph.

234. Furthermore, the smallest value that x.y can take on is 1.0. Values below 1.0 would be re-expressed in scientific notation in my discussion above by moving the decimal place and reducing the value k by one. Therefore, the largest deflation due to rounding that could occur in Potential Reach is $.04\overline{9}$ over 1.0, or simply $.04\overline{9}$, which is infinitely close to but smaller than 0.05 or 5%.

235. As this is a somewhat mathematical discussion, I offer examples in Appendix 3 to show that, regardless of the magnitude of the number, the maximum reduction in inflation due to rounding is less than 5%.

236. As discussed in Section VII above, rates of Potential Reach inflation that I calculate based on Facebook's own data are far greater than 5%. Moreover, Facebook's rounding of Potential Reach can result in inflation Potential Reach even more than it already is. Therefore, I did not consider Facebook's rounding in my calculations of Potential Reach inflation.